

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2006, 12:42:35 ; Search time 77.1431 Seconds
(without alignments)
1378.343 Million cell updates/sec

Title: US-10-635-977-2_COPY_133_374

Perfect score: 1299

Sequence: 1 IDGLNIIWIKPAKSRGRD.....IKVAVDRSCDIGNFELLWRQ 242

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_21.*

- 1: Geneseqp1980s.*
- 2: Geneseqp1990s.*
- 3: Geneseqp2000s.*
- 4: Geneseqp2001s.*
- 5: Geneseqp2002s.*
- 6: Geneseqp2003as.*
- 7: Geneseqp2003bs.*
- 8: Geneseqp2004s.*
- 9: Geneseqp2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1299	100.0	242	8	ADJ93360 Human BGS
2	1299	100.0	541	8	ADJ93365 Human tub
3	1299	100.0	541	8	ADJ93358 Human BGS
4	1247	96.0	293	8	ADJ93366 Human BGS
5	850	65.4	352	3	AAB43005 Human ORF
6	850	65.4	352	4	AAM39450 Human pol
7	850	65.4	352	8	ADJ93457 Human HOT
8	850	65.4	352	8	ABM80420 Tumour-as
9	850	65.4	362	3	AAB58909 Breast an
10	850	65.4	399	6	ABU11512 Human WDD
11	850	65.4	488	5	AAB74334 Human cyt
12	850	65.4	744	4	AAB94796 Human pro
13	847	65.2	326	7	ADM05524 Human pro
14	846	65.1	292	8	ADJ93455 Human HOT
15	628.5	48.4	992	4	ABB65645 Drosophil
16	555	42.7	432	6	ABU92048 Human pro
17	493.5	38.5	719	4	ABB65541 Drosophil
18	468	36.0	281	8	ADQ66614 Novel hum
19	393	30.3	362	4	AAM41236 Human pol
20	337	25.9	566	4	ABG05971 Novel hum
21	324.5	25.0	347	6	ABU00150 Human nov
22	324	24.9	496	4	ABB64074 Drosophil
23	297.5	22.9	827	4	ABB60840 Drosophil
24	296	22.8	1281	4	AAM39105 Human pol

ALIGNMENTS

RESULT 1

ADJ93360

ID ADJ93360 standard; protein; 242 AA.

XX AC ADJ93360;

XX DT 06-MAY-2004 (first entry)

XX DE Human BGS-42 protein sequence SeqID4.

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

Aam40891 Human pol
Abp43750 Ligase TT
Adj93458 Human tub
Abb60296 Drosophil
Adc33066 Human nov
Adc31411 Human nov
Adq66481 Novel hum
Abb59096 Drosophil
Adj93373 Human BGS
Adb65504 Human pro
Aam79305 Human pro
Abb22915 Protein #
Adb65217 Human pro
Abb64837 Drosophil
Adj71954 Human PMW
Adh45424 Human mol
Adj93456 Pig tubul
Aam79068 Human pro
Abb83472 Human cyt
Ade47756 Human NOV
Adj79026 Human NOV

testis-specific tubulin tyrosine-ligase-like polypeptide;
BGS-42 polypeptide; cytotstatic; respiratory-Gen; gastrointestinal-Gen;
neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;
osteopathic; nootropic; antiparkinsonian; antiarthritic; antiasthmatic;
anti-HIV; antibacterial; immunosuppressive; antiseborrheic;
dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;
tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;
testicular disorder; testicular cancer; pulmonary disorder; lung cancer;
gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;
brain cancer; liver cancer; proliferative condition; testis; lung;
small intestine; brain; lymph tissue; infertility; Cushing's syndrome;
emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;
Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;
sepsis; acne; Sjogren's disease; scleroderma; human.

New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,
useful for preventing, treating or ameliorating a medical condition, e.g.
aberrant cellular proliferation, reproductive disorders or testicular
disorders.

Disclosure: SEQ ID NO 4; 343pp; English.

XX This invention relates to a novel testis-specific tubulin tyrosine-ligase
 CC -like polypeptide, designated the BGS-42 polypeptide. The invention may
 CC be useful for the development of compounds with a cytostatic, respiratory
 CC -Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,
 CC antiinflammatory, anabolic, hypertensive, osteopathic, nootropic,
 CC antiparkinsonian, antiarthritic, antiasthmatic, anti-HIV, antibacterial,
 CC immunosuppressive, anticancer or dermatological activity acting as
 CC tyrosine ligase modulators. In addition, the disclosed sequences may be
 CC useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be
 CC used for diagnosing a pathological condition or a susceptibility to a
 CC pathological condition in a subject, and for preventing, treating or
 CC ameliorating a medical condition, such as a disorder related to aberrant
 CC tubulin ligase activity, a disorder related to aberrant tubulin-
 CC carboxypeptidase activity, aberrant cellular proliferation, reproductive
 CC disorders, testicular disorders, testicular cancer, pulmonary disorders,
 CC lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
 CC neural disorders, brain cancer, liver cancer, or proliferative condition
 CC of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
 CC polypeptide, polynucleotide, or their modulators are also useful for
 CC treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
 CC disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
 CC -42 polypeptide can be used as a preventive agent for immunological
 CC disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
 CC disease or scleroderma. The antibodies may be used to purify, detect and
 CC target the BGS-42 polypeptides. The present sequence is that of the human
 CC BGS-42 protein (partial sequence) of the invention.
 XX
 SQ Sequence 242 AA;

Query Match 100.0%; Score 1299; DB 8; Length 242;
 Best Local Similarity 100.0%; Pred. No. 1.9e-134;
 Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 IDGLRNIWIKPAKSRGRDIVCMRVRETELELAADHPLSRDNKVVQKYIETPLICD 60
 Db 1 IDGLRNIWIKPAKSRGRDIVCMRVRETELELAADHPLSRDNKVVQKYIETPLICD 60
 Qy 61 TKFDIRQWFLVTDWNPITWIFKESYLRFSTQFSLDKLDSAIHLCCNNAVQKYLKNDVGR 120
 Db 61 TKFDIRQWFLVTDWNPITWIFKESYLRFSTQFSLDKLDSAIHLCCNNAVQKYLKNDVGR 120
 Qy 121 SPILLPAHNMTSTRFOEYLQQRGAVGWSVIYPSMKKATAHAMKVAQDHVEPRKNSFEL 180
 Db 121 SPILLPAHNMTSTRFOEYLQQRGAVGWSVIYPSMKKATAHAMKVAQDHVEPRKNSFEL 180
 Qy 181 YGADFVLGRDFRPLWLEINSPTHPSTPTVTAQLCAQVQEDTIKVAVDRCSDIGNFELLW 240
 Db 181 YGADFVLGRDFRPLWLEINSPTHPSTPTVTAQLCAQVQEDTIKVAVDRCSDIGNFELLW 240
 Qy 241 RQ 242
 Db 241 RQ 242

RESULT 2
 ADJ93365
 ID ADJ93365 standard; protein; 541 AA.

XX ADJ93365;

XX 06-MAY-2004 (first entry)

XX Human tubulin tyrosine ligase protein consensus sequence SeqID13.

XX testis-specific tubulin tyrosine-ligase-like polypeptide;
 KW BGS-42 polypeptide; cytostatic; respiratory-Gen; Gastrointestinal-Gen;
 KW neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;
 KW osteopathic; nootropic; antiparkinsonian; antiarthritic; antiasthmatic;
 KW anti-HIV; antibacterial; immunosuppressive; anticancer;
 KW dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;
 KW tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;
 KW testicular disorder; testicular cancer; pulmonary disorder; lung cancer;

KW gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;
 KW brain cancer; liver cancer; proliferative condition; testis; lung;
 KW small intestine; brain; lymph tissue; infertility; Cushing's syndrome;
 KW emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;
 KW Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;
 KW sepsis; acne; Sjogren's disease; scleroderma; human.
 XX
 XX Homo sapiens.
 OS Synthetic.
 XX
 XX WO2004005487-A2.
 PN
 PD 15-JAN-2004.
 XX
 XX 09-JUL-2003; 2003WO-US021605.
 PF
 XX 09-JUL-2002; 2002US-0394725P.
 PR
 XX (BRIM) BRISTOL-MYERS SQUIBB CO.
 PA
 XX Feder JN, Wu S, Nelson TC;
 PI
 XX WPI; 2004-099381/10.
 DR
 XX N-PSDB; ADJ93364.
 XX
 PT New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,
 PT useful for preventing, treating or ameliorating a medical condition, e.g.
 PT aberrant cellular proliferation, reproductive disorders or testicular
 PT disorders.
 XX
 XX Example 4; SEQ ID NO 13; 343pp; English.

XX This invention relates to a novel testis-specific tubulin tyrosine-ligase
 CC -like polypeptide, designated the BGS-42 polypeptide. The invention may
 CC be useful for the development of compounds with a cytostatic, respiratory
 CC -Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,
 CC antiinflammatory, anabolic, hypertensive, osteopathic, nootropic,
 CC antiparkinsonian, antiarthritic, antiasthmatic, anti-HIV, antibacterial,
 CC immunosuppressive, anticancer or dermatological activity acting as
 CC tyrosine ligase modulators. In addition, the disclosed sequences may be
 CC useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be
 CC used for diagnosing a pathological condition or a susceptibility to a
 CC pathological condition in a subject, and for preventing, treating or
 CC ameliorating a medical condition, such as a disorder related to aberrant
 CC tubulin ligase activity, a disorder related to aberrant tubulin-
 CC carboxypeptidase activity, aberrant cellular proliferation, reproductive
 CC disorders, testicular disorders, testicular cancer, pulmonary disorders,
 CC lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
 CC neural disorders, brain cancer, liver cancer, or proliferative condition
 CC of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
 CC polypeptide, polynucleotide, or their modulators are also useful for
 CC treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
 CC disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
 CC -42 polypeptide can be used as a preventive agent for immunological
 CC disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
 CC disease or scleroderma. The antibodies may be used to purify, detect and
 CC target the BGS-42 polypeptides. The present sequence is that of the
 CC tubulin tyrosine ligase protein consensus sequence which was used in the
 CC exemplification of the invention.

XX Sequence 541 AA;

Query Match 100.0%; Score 1299; DB 8; Length 541;
 Best Local Similarity 100.0%; Pred. No. 6e-134;
 Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IDGLRNIWIKPAKSRGRDIVCMRVRETELELAADHPLSRDNKVVQKYIETPLICD 60
 Db 133 IDGLRNIWIKPAKSRGRDIVCMRVRETELELAADHPLSRDNKVVQKYIETPLICD 192
 Qy 61 TKFDIRQWFLVTDWNPITWIFKESYLRFSTQFSLDKLDSAIHLCCNNAVQKYLKNDVGR 120
 Db 193 TKFDIRQWFLVTDWNPITWIFKESYLRFSTQFSLDKLDSAIHLCCNNAVQKYLKNDVGR 252

QY 121 SPLLPAHNMTSTRFOEYLQRCRGAVMGSVIYPSMKKAIAHAMKVAQDHVEPRKNSFEL 180
 DB 253 SPLLPAHNMTSTRFOEYLQRCRGAVMGSVIYPSMKKAIAHAMKVAQDHVEPRKNSFEL 312
 QY 181 YGADFVLGRDPRFWLIEINSSPTMHPSTPTAQLCAQVQEDTIKVAVDRSCDIGNFELW 240
 DB 313 YGADFVLGRDPRFWLIEINSSPTMHPSTPTAQLCAQVQEDTIKVAVDRSCDIGNFELW 372
 QY 241 RQ 242
 DB 373 RQ 374

RESULT 3
 ID ADJ93358
 XX ADJ93358
 AC ADJ93358;
 XX
 DT 06-MAY-2004 (first entry)
 DE
 DE Human BGS-42 protein sequence SeqID2.
 XX
 KW testis-specific tubulin tyrosine-ligase-like polypeptide;
 KW BGS-42 polypeptide; cytosolic; respiratory-Gen; gastrointestinal-Gen;
 KW neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;
 KW osteopathic; nootropic; antiparkinsonian; antiarthritic; antiasthmatic;
 KW anti-HIV; antibacterial; immunosuppressive; antiseborrheic;
 KW dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;
 KW tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;
 KW testicular disorder; testicular cancer; pulmonary disorder; lung cancer;
 KW gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;
 KW brain cancer; liver cancer; proliferative condition; testis; lung;
 KW small intestine; brain; lymph tissue; infertility; Cushing's syndrome;
 KW emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;
 KW Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;
 KW sepsis; acne; Sjogren's disease; scleroderma; human.
 XX
 OS Homo sapiens.
 XX
 PN WO2004005487-A2.
 XX
 PD 15-JAN-2004.
 XX
 PF 09-JUL-2003; 2003WO-US021605.
 XX
 PR 09-JUL-2002; 2002US-0394725P.
 XX
 PA (BRIM) BRISTOL-MYERS SQUIBB CO.
 XX
 PI Feder JN, Wu S, Nelson TC;
 XX
 DR WPI; 2004-099381/10.
 DR N-FSDB; ADJ93357.
 XX
 XX New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,
 PT useful for preventing, treating or ameliorating a medical condition, e.g.
 PT aberrant cellular proliferation, reproductive disorders or testicular
 PT disorders.
 XX
 PS Claim 5; SEQ ID NO 2; 343pp; English.
 XX
 CC This invention relates to a novel testis-specific tubulin tyrosine-ligase
 CC -like polypeptide, designated the BGS-42 polypeptide. The invention may
 CC be useful for the development of compounds with a cytostatic, respiratory
 CC -Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,
 CC antiinflammatory, anabolic, hypertensive, osteopathic, nootropic,
 CC antiparkinsonian, antiarthritic, antiasthmatic, anti-HIV, antibacterial,
 CC immunosuppressive, antiseborrheic or dermatological activity acting as
 CC tyrosine ligase modulators. In addition, the disclosed sequences may be
 CC useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be
 CC used for diagnosing a pathological condition or a susceptibility to a

CC pathological condition in a subject, and for preventing, treating or
 CC ameliorating a medical condition, such as a disorder related to aberrant
 CC tubulin ligase activity, a disorder related to aberrant tubulin-
 CC carboxypeptidase activity, aberrant cellular proliferation, reproductive
 CC disorders, testicular disorders, testicular cancer, pulmonary disorders,
 CC lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
 CC neural disorders, brain cancer, liver cancer, or proliferative condition
 CC of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
 CC polypeptide, polynucleotide, or their modulators are also useful for
 CC treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
 CC disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
 CC -42 polypeptide can be used as a preventive agent for immunological
 CC disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
 CC disease or scleroderma. The antibodies may be used to purify, detect and
 CC target the BGS-42 polypeptides. The present sequence is that of the human
 CC BGS-42 protein of the invention.
 XX
 SQ Sequence 541 AA;

Query Match 100.0%; Score 1299; DB 8; Length 541;
 Best Local Similarity 100.0%; Pred. No. 6e-134;
 Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IDGLRNIIWIKPAKSRGRDIVCMRVVEILELAADHPLSRDNKVVQKYIETPLLICD 60
 DB 133 IDGLRNIIWIKPAKSRGRDIVCMRVVEILELAADHPLSRDNKVVQKYIETPLLICD 192
 QY 61 TKEDIRQWFLVTDWNPITWFKESYLRFTQRFSLDKLDSAIHLCNNAVQKYLKNDVGR 120
 DB 193 TKEDIRQWFLVTDWNPITWFKESYLRFTQRFSLDKLDSAIHLCNNAVQKYLKNDVGR 252
 QY 121 SPLPAHNMTSTRFOEYLQRCRGAVMGSVIYPSMKKAIAHAMKVAQDHVEPRKNSFEL 180
 DB 253 SPLPAHNMTSTRFOEYLQRCRGAVMGSVIYPSMKKAIAHAMKVAQDHVEPRKNSFEL 312
 QY 181 YGADFVLGRDPRFWLIEINSSPTMHPSTPTAQLCAQVQEDTIKVAVDRSCDIGNFELW 240
 DB 313 YGADFVLGRDPRFWLIEINSSPTMHPSTPTAQLCAQVQEDTIKVAVDRSCDIGNFELW 372
 QY 241 RQ 242
 DB 373 RQ 374

RESULT 4
 ID ADJ93366
 XX ADJ93366
 AC ADJ93366;
 XX
 DT 06-MAY-2004 (first entry)
 DE
 DE Human BGS-42 protein-related TTL1 domain.
 XX
 KW testis-specific tubulin tyrosine-ligase-like polypeptide;
 KW BGS-42 polypeptide; cytosolic; respiratory-Gen; gastrointestinal-Gen;
 KW neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;
 KW osteopathic; nootropic; antiparkinsonian; antiarthritic; antiasthmatic;
 KW anti-HIV; antibacterial; immunosuppressive; antiseborrheic;
 KW dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;
 KW tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;
 KW testicular disorder; testicular cancer; pulmonary disorder; lung cancer;
 KW gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;
 KW brain cancer; liver cancer; proliferative condition; testis; lung;
 KW small intestine; brain; lymph tissue; infertility; Cushing's syndrome;
 KW emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;
 KW Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;
 KW sepsis; acne; Sjogren's disease; scleroderma; human; TTL1 domain.
 XX
 OS Homo sapiens.
 XX
 PN WO2004005487-A2.
 XX
 PD 06-MAY-2004 (first entry)
 XX
 PF 09-JUL-2003; 2003WO-US021605.
 XX
 PR 09-JUL-2002; 2002US-0394725P.
 XX
 PA (BRIM) BRISTOL-MYERS SQUIBB CO.
 XX
 PI Feder JN, Wu S, Nelson TC;
 XX
 DR WPI; 2004-099381/10.
 DR N-FSDB; ADJ93357.
 XX
 XX New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,
 PT useful for preventing, treating or ameliorating a medical condition, e.g.
 PT aberrant cellular proliferation, reproductive disorders or testicular
 PT disorders.
 XX
 PS Claim 5; SEQ ID NO 2; 343pp; English.
 XX
 CC This invention relates to a novel testis-specific tubulin tyrosine-ligase
 CC -like polypeptide, designated the BGS-42 polypeptide. The invention may
 CC be useful for the development of compounds with a cytostatic, respiratory
 CC -Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,
 CC antiinflammatory, anabolic, hypertensive, osteopathic, nootropic,
 CC antiparkinsonian, antiarthritic, antiasthmatic, anti-HIV, antibacterial,
 CC immunosuppressive, antiseborrheic or dermatological activity acting as
 CC tyrosine ligase modulators. In addition, the disclosed sequences may be
 CC useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be
 CC used for diagnosing a pathological condition or a susceptibility to a

PD 15-JAN-2004.
XX
PF 09-JUL-2003; 2003WO-US021605.
XX
PR 09-JUL-2002; 2002US-0394725P.
XX
PA (BRIM) BRISTOL-MYERS SQUIBB CO.
XX
PI Feder JN, Wu S, Nelson TC;
XX
DR WPI; 2004-099381/10.
XX
XX New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,
PT useful for preventing, treating or ameliorating a medical condition, e.g.
PT aberrant cellular proliferation, reproductive disorders or testicular
PT disorders.
XX
XX Disclosure; SEQ ID NO 14; 343pp; English.
XX
CC This invention relates to a novel testis-specific tubulin tyrosine-ligase
CC -like polypeptide, designated the BGS-42 polypeptide. The invention may
CC be useful for the development of compounds with a cytostatic, respiratory
CC -Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,
CC antiinflammatory, anabolic, hypertensive, osteopathic, nootropic,
CC antiparkinsonian, antiarthritic, antiaesthatic, anti-HIV, antibacterial,
CC immunosuppressive, antiseborrheic or dermatological activity acting as
CC tyrosine ligase modulators. In addition, the disclosed sequences may be
CC useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be
CC used for diagnosing a pathological condition or a susceptibility to a
CC pathological condition in a subject, and for preventing, treating or
CC ameliorating a medical condition, such as a disorder related to aberrant
CC tubulin ligase activity, a disorder related to aberrant tubulin-
CC carboxypeptidase activity, aberrant cellular proliferation, reproductive
CC disorders, testicular disorders, testicular cancer, pulmonary disorders,
CC lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
CC neural disorders, brain cancer, liver cancer, or proliferative condition
CC of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
CC polypeptide, polynucleotide, or their modulators are also useful for
CC treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
CC disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
CC -42 polypeptide can be used as a preventive agent for immunological
CC disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
CC disease or scleroderma. The antibodies may be used to purify, detect and
CC target the BGS-42 polypeptides. The present sequence is that of the TIL
CC domain of the human BGS-42 protein of the invention.
XX
XX Sequence 293 AA;

Query Match 96.0%; Score 1247; DB 8; Length 293;
Best Local Similarity 100.0%; Pred. No. 1.4e-128;
Matches 233; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 IDGLRNWIITKPAKSRGRDVCMDRVEEILELAADHPLSRDKWVQVQYIETPLIICD 60
DB 61 IDGLRNWIITKPAKSRGRDVCMDRVEEILELAADHPLSRDKWVQVQYIETPLIICD 120
QY 61 TKFDIRQWFLVTDWNPNTIWFYKESYLRFSTQRFSLDKLDSAIHLNNVQVYKLVNDVGR 120
DB 121 TKFDIRQWFLVTDWNPNTIWFYKESYLRFSTQRFSLDKLDSAIHLNNVQVYKLVNDVGR 180
QY 121 SPLLPAINMTSTRFQYLOQRGAVGWSVIYPSMKKATAHAKVQADHVEPRKNSFEL 180
DB 181 SPLLPAINMTSTRFQYLOQRGAVGWSVIYPSMKKATAHAKVQADHVEPRKNSFEL 240
QY 181 YGADFVLGRDPRPWLIEINSSPTWHPSTPTVTAQLCAQVQEDTIKVAVDRSCDI 233
DB 241 YGADFVLGRDPRPWLIEINSSPTWHPSTPTVTAQLCAQVQEDTIKVAVDRSCDI 293

RESULT 5
AAB43005
ID AAB43005 standard; protein; 352 AA.
XX

AC AAB43005;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORFX ORF2769 polypeptide sequence SEQ ID NO:5538.
XX
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnery; antiposoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cholesterol ester storage; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive.
XX
XX Homo sapiens.
OS
XX WO200058473-A2.
PN
XX 05-OCT-2000.
PD
XX 31-MAR-2000; 2000WO-US008621.
PF
XX 31-MAR-1999; 99US-0127607P.
PR 02-APR-1999; 99US-0127636P.
PR 05-APR-1999; 99US-0127728P.
PR 30-MAR-2000; 2000US-00540763.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Shimkets RA, Leach M;
PI
XX WPI; 2000-602362/57.
DR N-PSDB; AAC77214.
XX
XX Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease.
XX
XX Claim 11; Page 4720-4721; 5507pp; English.
XX
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
CC antiposoriatic; antiparkinsonian; nootropic; neuroprotective; osteopathic;
CC anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;
CC cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
CC dermatological; immunosuppressive; antinflammatory; antibacterial;
CC antiviral; antifungal; antirheumatic; antithyroid; and antianemic. The
CC sequences can be used for determining the presence of or predisposition
CC to, or preventing or treating pathological conditions associated with an
CC ORFX-associated disorder. The nucleic acids can be used to express ORFX
CC proteins in gene therapy vectors. The proteins and nucleic acids may be
CC used to treat cancers, proliferative disorders, neurodegenerative
CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
CC storage, systemic lupus erythematosus, severe combined immunodeficiency
CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
CC cartilage damage, nocturnal haemoglobinuria, antinflammatory disease; to
CC enhance coagulation; to inhibit thrombosis; and as a contraceptive
XX
XX Sequence 352 AA;

Query Match 65.4%; Score 850; DB 3; Length 352;
Best Local Similarity 61.8%; Pred. No. 1.2e-84;
Matches 152; Conservative 41; Mismatches 49; Indels 4; Gaps 1;

XX	09-JUL-2002; 2002US-0394725P.
PR	(BRIM) BRISTOL-MYERS SQUIBB CO.
PA	Feder JN, Wu S, Nelson TC;
PI	WPI; 2004-099381/10.
DR	New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,
XX	useful for preventing, treating or ameliorating a medical condition, e.g.
PT	aberrant cellular proliferation, reproductive disorders or testicular
PT	disorders.
XX	Disclosure; SEQ ID NO 7; 343pp; English.
PS	This invention relates to a novel testis-specific tubulin tyrosine-ligase
XX	-like polypeptide, designated the BGS-42 polypeptide. The invention may
CC	be useful for the development of compounds with a cycostatic, respiratory
CC	-Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,
CC	antiinflammatory, anabolic, hypertensive, osteopathic, anti-HIV, antibacterial,
CC	antiparkinsonian, antiarthritic, antiasthmatic, immunosuppressive, aniseborrheic or dermatological activity acting as
CC	tyrosine ligase modulators. In addition, the disclosed sequences may be
CC	useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be
CC	used for diagnosing a pathological condition or a susceptibility to a
CC	pathological condition in a subject, and for preventing, treating or
CC	ameliorating a medical condition, such as a disorder related to aberrant
CC	tubulin ligase activity, a disorder related to aberrant tubulin-
CC	carboxypeptidase activity, aberrant cellular proliferation, reproductive
CC	disorders, testicular disorders, testicular cancer, pulmonary disorders,
CC	lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
CC	neural disorders, brain cancer, liver cancer, or proliferative condition
CC	of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
CC	polypeptide, polynucleotide, or their modulators are also useful for
CC	treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
CC	disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
CC	-42 polypeptide can be used as a preventive agent for immunological
CC	disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
CC	disease or scleroderma. The antibodies may be used to purify, detect and
CC	target the BGS-42 polypeptides. The present sequence is that of the human
CC	HOTTJ3 protein which is related to the invention. Note: The present
CC	sequence does not appear in the specification but was obtained from
CC	Genbank.
XX	
SQ	Sequence 352 AA;
	Query Match 65.4%; Score 850; DB 8; Length 352;
	Best Local Similarity 61.8%; Pred. No. 1.2e-84;
	Matches 152; Conservative 41; Mismatches 49; Indels 4; Gaps 1
Qy	1 IDGLRNITWIKPAKSRGRDTCVMDRVEEIIELAAADHPLSDKNKKVVQKYIETPLLICD 60
	MEGRDNTWIWKPKAKSRGRGIMCNDMHUEMLUKLVNGNPVMKGDKWVQKYIERPLLIIFG 120
Dd	61 TKFDIRQWLFTVDWNPLTIWFYKESYLRFSTORFSLDKLDSAIHLCNNAVOXYLKNDVGR 120
Qy	121 TKFDLRQWLFTVDWNPLTVFYRSYIRFSTQPFSKLNDNSVHLCKNNISIQHLENSCHR 180
Dd	121 SPLLPANHNMTSTRFBQYLQRQGRGAVMGSVIYPFSMKKAIAHAMKVQAHDHPEPRKNSFEL 180
Qy	181 HPLLPPDNMSSQRFQALHQEMGAPNAWSTIIVPGMKDAVIHALQTSDTQCRCASFEL 240
Dd	181 YGADFVLGRDPRPWLIENSPTWHPSTPTVAQLCAQVEDTIKAV-----DRSCDIGNF 236
Qy	241 YGADFVFEGFPQPLIEINASPTMAPSTAVTARLCAGVQADTLRVVIDMLDRNCDTGAF 300
Dd	237 ELLWRQ 242
Qy	301 ELIYKQ 306
Dd	
RESULT 8	

Db 121 TKFDLRQWFLVTDNNPLTWFYRDSIRSTQPSLKNDNSVHLNCNNSIQHLENSCHR 180
 QY 121 SPLLPAHNNMTSTRFOEYLQROGRGAVMGSVIYPSMKKAIAHAMKVAQDHVPRKNSFEL 180
 Db 181 HPLLPPDNWSSQRFQAHLEMGAPNASTIIVPGMKDAVIHALQTSQDTVQCRKASFEL 240
 QY 181 YGADFVLGRDPRFWLIEINSSPTMHPTVTAQLCAQVQEDTIKVAV-----DRSCDIGNF 236
 Db 241 YGADFVFGEDFQFWLIEINASTMAPSTAVTARLCAGVQADTLRVVIDRMLDRNCDTGAF 300
 QY 237 ELLWRQ 242
 Db 301 ELIYKQ 306

RESULT 9

AAB58909

ID AAB58909 standard; protein; 362 AA.

XX AAB58909;

DT 27-MAR-2001 (first entry)

DE Breast and ovarian cancer associated antigen protein sequence SEQ ID 617.

XX Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;
 KW neurotropic; neurprotection; antiviral; antiallergic; hepatotropic;
 KW antidiabetic; antinflammatory; antitumor; antitumor; antitumor;
 KW antibacterial; antifungal; antiparasitic; cardiac; immune disorder;
 KW Addison's disease; allergy; autoimmune haemolytic anaemia;
 KW autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
 KW multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
 KW cardiovascular disorder; wound healing; neurological disease.

XX Homo sapiens.

XX WO200055173-A1.

XX PD 21-SEP-2000.

XX PF 08-MAR-2000; 2000WO-US005881.

XX PR 12-MAR-1999; 99US-0124270P.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Rosen CA, Ruben SM;

XX WPI; 2000-611515/58.

XX DR N-FSDB; AAF21812.

XX New human breast and ovarian cancer associated gene sequences and the
 PT polypeptides encoded by these genes, useful in the prevention, treatment
 PT and diagnosis of cancer, immune disorders, cardiovascular disorders and
 PT neurological diseases.

XX PS Claim 11; Page 1056-1057; 1299pp; English.

XX Sequences AAF21614 - AAF22031 represent DNA sequences encoding human
 CC proteins AAB58911 - AAB59128. The DNA and protein sequences are
 CC associated with breast and ovarian cancer. Included in the invention are
 CC sequences AAF22032 - AAF22040 and AAB59129 which are used in the
 CC isolation and characterisation of the DNA and protein sequences of the
 CC invention. The breast and ovarian cancer associated DNA, protein, agonist
 CC or antagonist sequences exhibit cytostatic; immunosuppressive; neurotropic;
 CC neuroprotective; antiviral; antiallergic; hepatotropic; antidiabetic;
 CC antinflammatory; antitumor; antitumor; antitumor; antitumor;
 CC antifungal; antiparasitic and cardiac activity. The polynucleotide and
 CC protein sequences are used in the diagnosis of cancer, particularly
 CC breast and ovarian cancer. The nucleic acid sequences, proteins, agonists
 CC and agonists may also be used in the diagnosis, prevention and treatment
 CC of immune disorders e.g. Addison's disease, allergies, autoimmune

CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis;
 CC cardiovascular disorders such as myocardial ischaemia; wound healing;
 CC neurological diseases such as cerebral anoxia and epilepsy; and
 CC infectious diseases

XX Sequence 362 AA;

Query Match

Best Local Similarity 65.4%; Score 850; DB 3; Length 362;

Matches 152; Conservative 41; Mismatches 49; Indels 4; Gaps 1;

- QY 1 IDGLRNIIWIKPAKSGRDIIVCMRVRBEILELAADHPLSRDNKVVQKIETPLLICD 60
 :::::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 78 MEGDRNIWIKPGAKSGRGIMCMDBLEMLKLVNGNPVVMKDGKVVQKIETPLIFG 137
 QY 61 TKFDLRQWFLVTDNNPLTWFYRDSIRSTQPSLKNDNSVHLNCNNSIQHLENSCHR 120
 :::::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 138 TKFDLRQWFLVTDNNPLTWFYRDSIRSTQPSLKNDNSVHLNCNNSIQHLENSCHR 197
 QY 121 SPLLPAHNNMTSTRFOEYLQROGRGAVMGSVIYPSMKKAIAHAMKVAQDHVPRKNSFEL 180
 :::::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 198 HPLLPPDNWSSQRFQAHLEMGAPNASTIIVPGMKDAVIHALQTSQDTVQCRKASFEL 257
 QY 181 YGADFVLGRDPRFWLIEINSSPTMHPTVTAQLCAQVQEDTIKVAV-----DRSCDIGNF 236
 :::::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 258 YGADFVFGEDFQFWLIEINASTMAPSTAVTARLCAGVQADTLRVVIDRMLDRNCDTGAF 317
 QY 237 ELLWRQ 242
 :::::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 318 ELIYKQ 323

RESULT 10

ABU11512

ID ABU11512 standard; protein; 399 AA.

XX AC ABU11512;

XX DT 12-FEB-2003 (first entry)

XX DE Human MDDT polypeptide SEQ ID 459.

XX MDDT; human; disease detection and treatment molecule polypeptide;
 KW anti-inflammatory; immunosuppressive; osteopathic; cytostatic; anti-HIV;
 KW haemostatic; nephrotropic; antianaemic; anipsoriatic; hepatotropic;
 KW gene therapy; protein replacement therapy; cell proliferative disorder;
 KW cancer; adenocarcinoma; leukaemia; lymphoma; melanoma; myeloma; sarcoma;
 KW anaemia; Crohn's disease; acquired immunodeficiency syndrome; AIDS;
 KW Goodpasture's syndrome; inflammation; osteoporosis; thrombocytopaenia;
 KW psoriasis; hepatitis.

XX OS Homo sapiens.

XX XX WO200279449-A2.

XX PD 10-OCT-2002.

XX PF 27-MAR-2002; 2002WO-US009944.

XX PR 28-MAR-2001; 2001US-0279619P.

XX PR 29-MAR-2001; 2001US-0280067P.

XX PR 29-MAR-2001; 2001US-0280068P.

XX PR 16-MAY-2001; 2001US-0291280P.

XX PR 17-MAY-2001; 2001US-0291829P.

XX PR 19-JUN-2001; 2001US-0291849P.

XX PR 20-JUN-2001; 2001US-0299776P.

XX PR 20-JUN-2001; 2001US-0300001P.

XX PA (INCY-) INCYTE GENOMICS INC.

XX PI Dafo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;

QY 181 YGADFVLGRDRPRLWLEINSSPTMHPSTPTAQLCAQVQEDTIKAV-----DRSCDIGNP 236
DB 241 YGADFVFGEDFQPLWLEINASPTMAPSTAVTARLCAGVQADTLRVVVIDRMLDRNCDTGAF 300
QY 237 ELLWRQ 242
DB 301 ELIYKQ 306
RESULT 12
ID AAB94796
XX AAB94796 standard; protein; 744 AA.
AC AAB94796;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:15921.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-00116126.
XX
PR 29-JUL-1999; 99JP-00248036.
PR 27-AUG-1999; 99JP-00300253.
PR 11-JAN-2000; 2000JP-00187776.
PR 02-MAY-2000; 2000JP-00183767.
PR 09-JUN-2000; 2000JP-00241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.
XX
PS Claim 8; SEQ ID NO 15921; 2537pp + Sequence Listing; English.
XX
CC The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention
XX
SQ Sequence 744 AA;

Query Match 65.4%; Score 850; DB 4; Length 744;
Best Local Similarity 61.8%; Pred. No. 3.7e-84;
Matches 152; Conservative 41; Mismatches 49; Indels 4; Gaps 1;
QY 1 IDGLRNITIKPAKSRGRDIVCMRVERBILSLAADHPLSRDNKVVVKYIETPLLICD 60
DB 273 MEGDRNITWIKPKAKSRGRGIMCMHLEMLKLVNGNPVVMKDGKVVVKYIERPILIFG 332
QY 61 TKFDIRQWFLVTDNPLTIWFYKESYLFSTQRFSLDKLSDSAIHLCNNAVQKYLKNDVGR 120
DB 333 TKFDLRQWFLVTDNPLTIWFYRDSYIRFSTQRFSLKNDNSVHLNCNNSIQHLENSCHR 392
QY 121 SPLPAHNMWTSRFEYLQROGRGAVMGSVIYPSMKKAIAHAMKVAQDHVPRKNSFEL 180
DB 393 HELLPPDNWSSORFOAHLQEMGAPNAWSTIIVPGMKDAVIHALQTSQDTVQCRKASFEL 452
QY 181 YGADFVLGRDRPRLWLEINSSPTMHPSTPTAQLCAQVQEDTIKAV-----DRSCDIGNP 236
DB 453 YGADFVFGEDFQPLWLEINASPTMAPSTAVTARLCAGVQADTLRVVVIDRMLDRNCDTGAF 512
QY 237 ELLWRQ 242
DB 513 ELIYKQ 518
RESULT 13
ADM05524
ID ADM05524 standard; protein; 326 AA.
XX
AC ADM05524;
XX
DT 20-MAY-2004 (first entry)
XX
DE Human protein of the invention SEQ ID NO:4209.
XX
KW human; gene therapy; diagnostic marker; pharmaceutical.
XX
OS Homo sapiens.
XX
PN EP1347046-A1.
XX
PD 24-SEP-2003.
XX
PF 12-APR-2002; 2002EP-00008400.
XX
PR 22-MAR-2002; 2002JP-00137785.
XX
PA (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI Seki N, Yoshikawa T, Otsuka M, Negahari K, Masuho Y;
XX
DR WPI; 2003-723558/69.
XX
N-PSDB; ADM03081.
XX
PT New polynucleotides and polypeptides are useful in gene therapy, for developing a diagnostic marker or medicines for regulating their expression and activity, or as a target of gene therapy.
XX
PS Claim 1; SEQ ID NO 4209; 305pp; English.
XX
CC The invention relates to a novel human polynucleotide and the encoded polypeptide. A polynucleotide of the invention may have a use in gene therapy. An oligonucleotide of the invention ADM06202-ADM06773 is useful as a primer for synthesizing the polynucleotide or as a probe for detecting the polynucleotide. The polynucleotides ADM01316-ADM03758 are useful in gene therapy, for developing a diagnostic marker or medicines for regulating their expression and activity, or as a target of gene therapy. The proteins ADM03759-ADM06201 encoded by the polynucleotides are useful as pharmaceutical agents. The present sequence represents a protein sequence of the invention.

XX WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US009231.
XX PR 23-MAR-2000; 2000US-0191637P.
XX PR 11-JUL-2000; 2000US-00614150.
XX PA (PEKE) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX DR WPI; 2001-656860/75.
XX DR N-PSDB; ABL09748.
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signaling and cell-cell
XX PT interactions.
XX PS Disclosure; SEQ ID NO 23727; 21pp + Sequence Listing; English.
XX PS The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-
XX CC ABB72072). The sequence data for this patent did not form part of the
XX CC printed specification, but was obtained in electronic format directly
XX CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 992 AA;

Query Match 48.4%; Score 628.5; DB 4; Length 992;
Best Local Similarity 48.4%; Pred. No. 1.6e-59;
Matches 119; Conservative 50; Mismatches 70; Indels 7; Gaps 3;

Qy 1 IDGLRNWIKPAKSRGRDIVCMRDVREILELAADHPLSRDNKWWQKYIETPLICD 60
Db 411 LDGYQNMWIVKPAKRCRGRGIILMDNLKII--LGVVNLISAKSRVYQVQYIERPLILFQ 468
Qy 61 TKFDIRQWFLVTDWNPETIWFYKESYLRFSTQRFSLDKLDSAIHLGNNAVQKYLKNDVGR 120
Db 469 TKFDIRQWFLIINTQPLVWVWFYRESYLRFSSQESYLSNHHESVHLTNTYAIQKKTNG-KR 527
Qy 121 SPFLPAHNNWTSRFOEYLQROGRGAVGSGVIYPSMKKAIAHAMKVAQDHVEPRKNSPEL 180
Db 528 DKRLPSENWDCYCSFOAYLRQIKYKNMMLERIFPGMRKAIVGCMLASQENMDRRPNTFEL 587
Qy 181 YGADFVLGRDPRFWLIEINSSPTMHPSTPTVTAQLCAQVOEDTIKYAVDRSCD---IGNF 236
Db 588 FGADFMCNCFPWLIEINSSPDLGATTSVTARMCPQCLEDVVKVVIDRRTDPKAEI LGNF 647
Qy 237 ELLWRQ 242
Db 648 ELAYRQ 653

Search completed: April 4, 2006, 12:51:06
Job time : 78.1431 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7

Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2006, 12:59:01 ; Search time 6.53532 Seconds
(without alignments)
1364.812 Million cell updates/sec

Title: US-10-635-977-2 COPY 73 365

Perfect score: 1560
Sequence: 1 ED1TSDAIVEDLTEAEWED.....LCAQVQEDTIKVAVDKSDI 293

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 180808 seqs, 30441898 residues

Total number of hits satisfying chosen parameters: 180808

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 s

Database : Published Applications AA New:*

- 1: $SIDS5^{\dagger}ptdata_2/2/pubpaa/US08$ NEW PUB pep:
- 2: $SIDS5^{\dagger}ptdata_2/2/pubpaa/US06$ NEW PUB pep:
- 3: $SIDS5^{\dagger}ptdata_2/2/pubpaa/US07$ NEW PUB pep:
- 4: $SIDS5^{\dagger}ptdata_2/pubpaa/PCT$ NEW PUB pep:
- 5: $SIDS5^{\dagger}ptdata_2/2/pubpaa/US09$ NEW PUB pep:
- 6: $SIDS5^{\dagger}ptdata_2/2/pubpaa/US10$ NEW PUB pep:
- 7: $SIDS5^{\dagger}ptdata_2/2/pubpaa/US11$ NEW PUB pep:
- 8: $SIDS5^{\dagger}ptdata_2/2/pubpaa/US60$ NEW PUB pep:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	App
1	271.5	17.4	592	7	US-11-072-512-3371	Sequence 3371, Ap	
2	256.5	16.4	439	7	US-11-072-513-3658	Sequence 3658, Ap	
3	180	11.5	404	7	US-11-072-513-3621	Sequence 3621, Ap	
4	93	6.0	550	7	US-11-096-568A-12196	Sequence 12196, A	
5	93	6.0	579	7	US-11-096-568A-12195	Sequence 12195, A	
6	93	6.0	596	7	US-11-096-568A-12194	Sequence 12194, A	
7	82	5.3	1137	7	US-11-087-099-5799	Sequence 5799, Ap	
8	80	5.1	729	6	US-10-995-561-878	Sequence 878, App	
9	80	5.1	744	6	US-10-995-561-876	Sequence 876, App	
10	80	5.1	752	6	US-10-995-561-879	Sequence 879, App	
11	80	5.1	753	6	US-10-995-561-877	Sequence 877, App	
12	79.5	5.1	302	7	US-11-098-686-10927	Sequence 10927, A	
13	79.5	5.1	433	7	US-11-167-856-26	Sequence 26, Appl	
14	79	5.1	462	7	US-11-096-568A-25994	Sequence 25994, A	
15	79	5.1	470	7	US-11-096-568A-25993	Sequence 25993, A	
16	79	5.1	517	7	US-11-096-568A-25992	Sequence 25992, A	
17	79	5.1	549	7	US-11-085-185-2	Sequence 2, Appli	
18	79	5.1	932	7	US-11-017-550-65	Sequence 65, Appl	
19	79	5.1	932	7	US-11-002-387-65	Sequence 65, Appl	
20	79	5.1	1131	7	US-11-087-099-2209	Sequence 2209, Ap	
21	79	5.1	1131	7	US-11-087-099-9522	Sequence 9522, Ap	
22	78	5.0	497	6	US-10-793-626-2812	Sequence 2812, Ap	
23	77	4.9	993	7	US-11-132-764-7	Sequence 7, Appli	
24	77	4.9	1010	6	US-10-959-611-2	Sequence 2, Appli	
25	76.5	4.9	127	7	US-11-087-099-2254	Sequence 2254, Ap	

Db 233 -FDMIVQKYNPLLIQRYKCDLRIYVCTGFKPLTIIVYQGLVRATEKFDLSNLQ 291
QY 160 DSAIHLNNAVOK-----YLNKDVGRSPLLPAHNMWTSRFOEYIQRQGRGAW---GS 210
Db 292 NYAHLTNSINKSGASYEKIKEVIGHG-----CKWTLRSRFFSYLRS-----WDVDDL 339
QY 211 VIYPSMKAATAHAMKVAQDHVEPRKNSFELYGADFVLGRDPRMLIEINSSP--TWHPS 268
Db 340 LLWKKTHRMVILTLAIPSPFAANCFLFGFDILIDDNLKPWLLVNVNYPALTLD 399
QY 269 PV 270
Db 400 DV 401

RESULT 2

US-11-072-512-3658
; Sequence 3658, Application US/11072512
; Publication No. US20060029945A1
; GENERAL INFORMATION:

; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOUYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: Novel full length cdNA
; FILE REFERENCE: 084335-0191
; CURRENT APPLICATION NUMBER: US/11/072,512
; CURRENT FILING DATE: 2005-03-07
; PRIOR FILING DATE: 2002-01-25
; PRIOR APPLICATION NUMBER: JP 2001-379298
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3658
; LENGTH: 439
; TYPE: PRT
; ORGANISM: Homo sapiens

US-11-072-512-3658

Query Match 16.4%; Score 256.5; DB 7; Length 439;
Best Local Similarity 25.5%; Pred. No. 2.1e-17;
Matches 84; Conservative 56; Mismatches 109; Indels 81; Gaps 13;

QY 10 VEDLTEAEWE-----DLT-----QOYYSLVHGDFAISNSRNYFSQC 45
Db 65 VEVKDEGEWFCYCDVSWLRENFDHTYNDHVRISHFRNHYELTRKNYMWNLKFRKQL 124
QY 46 QALINRITSVN---POT-----DIDGLRN-----IWIIPAKSRGRDIVCMRDRVE 88
Db 125 EREAGKLEAAKCDFFPKTFEMPCBYHLFVEBFKNPGITWIMKPVARSQKGIPLFRLLK 184
QY 89 EILEAADAHPLSRDNK-----WVQKIETPLLCIDTKFDIRQWFLVTDWNPPLIWP 141
Db 185 DIVDW-RKDRSSDDQKDDIPVENYVAQRVTENFYLIIGGRKFDLRVYLVMS-----V 236
QY 142 YKESYLFSTORFSLDKLSDSAIHLNNAVOKYLNKDVGRSPLLPAHN-----WMTSTRFQ 196

Db 237 FABELLWSGHR-----QDVHLTNVAVOK-----TSPDYHPKKGCKWTQRF 279
QY 197 EYL-QRQGRGAVGWSVIYPSMKAIAHAMKVAQDHVEPRKNSFELYGADFVLGRDPRPWL 255
Db 280 QYLASHKHGPEAV--ETLFRDIDNIFVKLSQSVQKVIISDKHCFELYGYDILIDODLKPWL 337
QY 256 IBNSSPPTWHPSPPTVTAQLCAQVOEDTIKV 285
Db 338 LEVNASPFLTASSQEDYELKTCLEDTLHV 367

RESULT 3

US-11-072-512-3621
; Sequence 3621, Application US/11072512
; Publication No. US20060029945A1
; GENERAL INFORMATION:

; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOUYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: Novel full length cdNA
; FILE REFERENCE: 084335-0191
; CURRENT APPLICATION NUMBER: US/11/072,512
; CURRENT FILING DATE: 2005-03-07
; PRIOR FILING DATE: 2002-01-25
; PRIOR APPLICATION NUMBER: JP 2001-379298
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3621
; LENGTH: 404
; TYPE: PRT
; ORGANISM: Homo sapiens

US-11-072-512-3621

Query Match 11.5%; Score 180; DB 7; Length 404;
Best Local Similarity 25.6%; Pred. No. 6.8e-10;
Matches 63; Conservative 33; Mismatches 84; Indels 66; Gaps 12;

QY 4 DTSADAVEDLTEAEWEDLTQOYYSLVHGDFAISNSRNYFSQCQALLNRITSVNPQT---D 60
Db 183 DAAAPAEDEL---PW-----TSPCYLAPQVRL--RMEFFETETRLD 219
QY 61 IDGLR-----NIWIIIPAKSRGRDIVCMRDRVEEILELAA-----ADHPLSRDNK 105
Db 220 LKHEREAFFTLFDETQIWKCTPATSNQKGIFFLLRNQEEVAALQAKTRSMEDDPHKKTP 279
QY 106 W-----VQKIETPLLCIDTKFDIRQWFLVTDWNPPLTIWFYKESYLRSTQFSLDKL 159
Db 280 FRGQARVQVQYIQNPLLVGRKFDVRSYLLIACTTPYMI--FFGHGYARLTLSLYDPHSS 338
QY 160 DSAIHLNNAVOKYLNKDVGRSP---LLPAHNMWTSRFOEYL-----QRQGRGAVGWSV 211
Db 339 DLGHLTNQFMOK-----KSPLYMLLKHTVWSMEHLNRYISDTFWKARGLAKDW--- 388
QY 212 IYPSMK 217
Db 389 VFTTLK 394


```
; Publication No. US20060041961A1
; GENERAL INFORMATION:
; APPLICANT: Abad, Mark S. et al.
; TITLE OF INVENTION: Genes and Uses for Plant Improvement
; FILE REFERENCE: 38-21(53450)B EP
; CURRENT APPLICATION NUMBER: US/11/087,099
; CURRENT FILING DATE: 2005-03-22
; NUMBER OF SEQ ID NOS: 12464
; SEQ ID NO 5799
; LENGTH: 1137
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1137)
; OTHER INFORMATION: unsure at all Xaa locations
US-11-087-099-5799

Query Match 5.1%; Score 82; DB 7; Length 1137;
Best Local Similarity 18.9%; Pred. No. 13;
Matches 41; Conservative 39; Mismatches 65; Indels 72; Gaps 8;

QY 43 SOQALLNRITSVNPQTDIDGLRNIIK-----PAKSRGRDIVCM-----DRVVE 89
Db 631 SEIVRLIDTATVILAVDVGVLNGWNKIAELTGLPIGEATGKHLTLTVEDSSDRVKK 690
QY 90 ILELAADHPLSRDNKRWVQKIETPLLCIDTKFDIRQWFLVTDWNPPLTIWFYKESYLR 149
Db 691 MLNLAL-----LGBEEK-----NVQFEIKTHGSKMDSGPISL-----V 723
QY 150 STQSFSLDKLDSAIHLC-----NNAVQYKLNK-----VGRSPLL----- 184
Db 724 VNACASRLDRNVGVCFVAHDITQAQKNVMDKRIEYDGYKAIQVNRNPLIPPIFGTDEF 783
QY 185 -----PAHNMWTSRFOYLQROGRGAVGWSVIYP 214
Db 784 GWCEWNPAMKLTGWKKEEVMKMLLGEIFGTQMAP 820

RESULT 8
US-10-995-561-878
; Sequence 878, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 878
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-995-561-878

Query Match 5.1%; Score 80; DB 6; Length 729;
Best Local Similarity 19.2%; Pred. No. 11;
Matches 51; Conservative 42; Mismatches 97; Indels 76; Gaps 11;

QY 43 SOQALLNRITSVNPQTDIDGLRNIIKPAKSRGRDIVCMRVSEILELAADHPLSR 102
Db 278 TDCENLLKRFVLNP-----IK-----RG-----TLEQIMK----- 303
QY 103 DNKRV-----VQKIETPLLCIDTK-----FDIRQWFLVTDWNPPLTIWFY 142
Db 304 -DRWINAGHEDELKPFVEPELDSQKRIDIMVGMGYSOEIQESLSKMKYDEIT---- 358
QY 143 KESYLRFSTQRFSLDKLDSA-----IHLCNNAVQYKLNKDVGRSPLLPAHNMWTSRFOEY 198
Db 359 -ATYLLGRKSSSELDASDSSSSNLSLAKVPSDDLNNSTGQSPHHKVQSVSSQKRR 417
QY 199 LQROGRGAVGWSVIYPSMKKAIHAHMKVAQDHVEPRKNSFELYCADFVLGRDFRPLWLEI 258
Db 418 YSDHAGPAIPSVVAYPKRSQTSADSLKEDGISSRKSSGSAVG-----GKGIAP----- 467
QY 259 NSSPMT-HPSTPVTQAOLCAQVQEDTI 283
Db 468 -ASPMGLGNASNPKNKADIPERKKSSTV 492

RESULT 10
US-10-995-561-879
; Sequence 879, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 879
; LENGTH: 752
; TYPE: PRT
US-10-995-561-879

Query Match 5.1%; Score 80; DB 6; Length 729;
Best Local Similarity 19.2%; Pred. No. 11;
Matches 51; Conservative 42; Mismatches 97; Indels 76; Gaps 11;

QY 43 SOQALLNRITSVNPQTDIDGLRNIIKPAKSRGRDIVCMRVSEILELAADHPLSR 102
Db 278 TDCENLLKRFVLNP-----IK-----RG-----TLEQIMK----- 303
QY 103 DNKRV-----VQKIETPLLCIDTK-----FDIRQWFLVTDWNPPLTIWFY 142
Db 304 -DRWINAGHEDELKPFVEPELDSQKRIDIMVGMGYSOEIQESLSKMKYDEIT---- 358
QY 143 KESYLRFSTQRFSLDKLDSA-----IHLCNNAVQYKLNKDVGRSPLLPAHNMWTSRFOEY 198
Db 418 YSDHAGPAIPSVVAYPKRSQTSADSLKEDGISSRKSSGSAVG-----GKGIAP----- 467
QY 259 NSSPMT-HPSTPVTQAOLCAQVQEDTI 283
Db 468 -ASPMGLGNASNPKNKADIPERKKSSTV 492
```


GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.
OM protein - protein search, using sw model
Run on: April 4, 2006, 12:55:16 ; Search time 42.2072 Seconds
(without alignments)
2900.544 Million cell updates/sec

Title: US-10-635-977-2_COPY_73_365
Perfect score: 1560
Sequence: 1 EDITSADAVEDETEAEWED.....LCAQVQEDTIKVAVDRSCDI 293

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA Main:
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1560	100.0	293	4	US-10-615-659-14
2	1560	100.0	293	4	US-10-635-977-14
3	1560	100.0	541	4	US-10-615-659-2
4	1560	100.0	541	4	US-10-615-659-13
5	1560	100.0	541	4	US-10-635-977-2
6	1560	100.0	541	4	US-10-635-977-13
7	1247	79.9	242	4	US-10-615-659-4
8	1247	79.9	242	4	US-10-635-977-4
9	908	58.2	352	4	US-10-615-659-7
10	908	58.2	352	4	US-10-635-977-7
11	908	58.2	352	5	US-10-756-149-5267
12	908	58.2	488	4	US-10-275-595A-5
13	905	58.0	326	4	US-10-108-260A-4209
14	897	57.5	362	3	US-09-923-298-617
15	897	57.5	362	4	US-10-102-806-617
16	817	52.4	292	4	US-10-615-659-5
17	817	52.4	292	4	US-10-635-977-5
18	636	40.8	992	6	US-11-097-143-23727
19	515.5	33.0	719	6	US-11-097-143-23415
20	374	24.0	566	5	US-10-450-763-36330
21	324	20.8	496	6	US-11-097-143-19014
22	321.5	20.6	330	4	US-10-615-659-26
23	321.5	20.6	330	4	US-10-635-977-26
24	297.5	19.1	827	6	US-11-097-143-9312
25	296	19.0	1226	5	US-10-756-149-5184
26	290.5	18.6	989	6	US-11-097-143-7680
27	287	18.4	423	4	US-10-615-659-8

28	287	18.4	423	4	US-10-635-977-8	Sequence 8, Appli
29	284	18.2	268	4	US-10-424-599-205823	Sequence 205823,
30	274	17.6	49	4	US-10-615-659-21	Sequence 21, Appl
31	274	17.6	49	4	US-10-615-659-22	Sequence 22, Appl
32	274	17.6	49	4	US-10-635-977-21	Sequence 21, Appl
33	274	17.6	49	4	US-10-635-977-22	Sequence 22, Appl
34	272	17.4	487	6	US-11-097-143-4080	Sequence 4080, Ap
35	271.5	17.4	524	3	US-09-864-761-38213	Sequence 38213, A
36	271.5	17.4	592	4	US-10-104-047-3371	Sequence 3371, Ap
37	270	17.3	917	6	US-11-097-143-21303	Sequence 21303, A
38	258	16.5	92	4	US-10-424-599-262294	Sequence 262294,
39	256.5	16.4	439	4	US-10-104-047-3658	Sequence 3658, Ap
40	255.5	16.4	379	4	US-10-615-659-6	Sequence 6, Appli
41	255.5	16.4	379	4	US-10-635-977-6	Sequence 6, Appli
42	250.5	16.1	377	4	US-10-210-130-118	Sequence 118, App
43	250.5	16.1	377	4	US-10-250-613-1	Sequence 1, Appli
44	208	13.3	720	6	US-11-097-143-12744	Sequence 12744, A
45	205.5	13.2	553	4	US-10-108-260A-2495	Sequence 2495, Ap

ALIGNMENTS

RESULT 1
US-10-615-659-14
; Sequence 14, Application US/10615659
; Publication No. US20040157234A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; CURRENT FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 14
; LENGTH: 293
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-615-659-14

Query Match	100.0%;	Score 1560;	DB 4;	Length 293;
Best Local Similarity	100.0%;	Pred. No. 2.7e-151;		
Matches 293;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	EDITSADAVEDETEAEWEDLTQQYSLVHGDAFISNRNYFSQCOALLNRITSVNPQTD	60	
Db	1	EDITSADAVEDETEAEWEDLTQQYSLVHGDAFISNRNYFSQCOALLNRITSVNPQTD	60	
Qy	61	IDGLRNIIIKPAKSRGRDIVCMRVVEIELEAAADHPLSRDNKWWVKYIETPLICD	120	
Db	61	IDGLRNIIIKPAKSRGRDIVCMRVVEIELEAAADHPLSRDNKWWVKYIETPLICD	120	
Qy	121	TKFDIROWFLVTDNPLTIWFKYSYLFSTQSFORSDLKLSAHLNNAVQKYLKNDVGR	180	
Db	121	TKFDIROWFLVTDNPLTIWFKYSYLFSTQSFORSDLKLSAHLNNAVQKYLKNDVGR	180	
Qy	181	SPLLPAHNMWTSRFOEYLRQGRGAVMGSVIYFSMKKAI AHAMKVAQDHVPRKNSFEL	240	
Db	181	SPLLPAHNMWTSRFOEYLRQGRGAVMGSVIYFSMKKAI AHAMKVAQDHVPRKNSFEL	240	
Qy	241	YGADFVLGRDRPRMLIEINSSPTMHPSTPTVAQLCAQVQEDTIKVAVDRSCDI	293	
Db	241	YGADFVLGRDRPRMLIEINSSPTMHPSTPTVAQLCAQVQEDTIKVAVDRSCDI	293	

RESULT 2
US-10-635-977-14
; Sequence 14, Application US/10635977
; Publication No. US20040171131A1

```
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635,977
; CURRENT FILING DATE: 2003-08-07
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S. 10/615,659
; PRIOR FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 14
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-635-977-14
```

```
Query Match      100.0%; Score 1560; DB 4; Length 293;
Best Local Similarity 100.0%; Pred. No. 2.7e-151;
Matches 293; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EDITSDADAVEDLTEAEWEDLTQQYYSLVHGDAPFINSRNYFSQCQALLNRITSVNPQTD 60
Db 1 EDITSDADAVEDLTEAEWEDLTQQYYSLVHGDAPFINSRNYFSQCQALLNRITSVNPQTD 60

Qy 61 IDGLRNIWIIKPAAKSRGRDIVCMDRVEEILEAAADHPLSRDNKWKVQKIETPLLICD 120
Db 61 IDGLRNIWIIKPAAKSRGRDIVCMDRVEEILEAAADHPLSRDNKWKVQKIETPLLICD 120

Qy 121 TKFDIRQWFLVTDWNPPLTIWFYKESYLRFSTQRFSLDKLDSAIHLCNNAVQKYLKNDVGR 180
Db 121 TKFDIRQWFLVTDWNPPLTIWFYKESYLRFSTQRFSLDKLDSAIHLCNNAVQKYLKNDVGR 180

Qy 181 SP LLPAHNMWTSRFOEYLQROGRGAVMGSVIYPSMKKAIAHAMKVAQDHVEPRKNSFEL 240
Db 181 SP LLPAHNMWTSRFOEYLQROGRGAVMGSVIYPSMKKAIAHAMKVAQDHVEPRKNSFEL 240

Qy 241 YGADFVLGRDPRFWLIEINSSPTMHPSTPVTQAQLCAQVQEDTIKVAVD RSCDI 293
Db 241 YGADFVLGRDPRFWLIEINSSPTMHPSTPVTQAQLCAQVQEDTIKVAVD RSCDI 293
```

```
RESULT 3
US-10-615-659-2
; Sequence 2, Application US/10615659
; Publication No. US20040157234A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; CURRENT FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-615-659-2

Query Match      100.0%; Score 1560; DB 4; Length 541;
Best Local Similarity 100.0%; Pred. No. 6.3e-151;
Matches 293; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EDITSDADAVEDLTEAEWEDLTQQYYSLVHGDAPFINSRNYFSQCQALLNRITSVNPQTD 60
Db 73 EDITSDADAVEDLTEAEWEDLTQQYYSLVHGDAPFINSRNYFSQCQALLNRITSVNPQTD 132
```

```
Qy 61 IDGLRNIWIIKPAAKSRGRDIVCMDRVEEILEAAADHPLSRDNKWKVQKIETPLLICD 120
Db 133 IDGLRNIWIIKPAAKSRGRDIVCMDRVEEILEAAADHPLSRDNKWKVQKIETPLLICD 192

Qy 121 TKFDIRQWFLVTDWNPPLTIWFYKESYLRFSTQRFSLDKLDSAIHLCNNAVQKYLKNDVGR 180
Db 193 TKFDIRQWFLVTDWNPPLTIWFYKESYLRFSTQRFSLDKLDSAIHLCNNAVQKYLKNDVGR 252

Qy 181 SP LLPAHNMWTSRFOEYLQROGRGAVMGSVIYPSMKKAIAHAMKVAQDHVEPRKNSFEL 240
Db 253 SP LLPAHNMWTSRFOEYLQROGRGAVMGSVIYPSMKKAIAHAMKVAQDHVEPRKNSFEL 312

Qy 241 YGADFVLGRDPRFWLIEINSSPTMHPSTPVTQAQLCAQVQEDTIKVAVD RSCDI 293
Db 313 YGADFVLGRDPRFWLIEINSSPTMHPSTPVTQAQLCAQVQEDTIKVAVD RSCDI 365
```

```
RESULT 4
US-10-615-659-13
; Sequence 13, Application US/10615659
; Publication No. US20040157234A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; CURRENT FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 13
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-615-659-13
```

```
Query Match      100.0%; Score 1560; DB 4; Length 541;
Best Local Similarity 100.0%; Pred. No. 6.3e-151;
Matches 293; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EDITSDADAVEDLTEAEWEDLTQQYYSLVHGDAPFINSRNYFSQCQALLNRITSVNPQTD 60
Db 73 EDITSDADAVEDLTEAEWEDLTQQYYSLVHGDAPFINSRNYFSQCQALLNRITSVNPQTD 132

Qy 61 IDGLRNIWIIKPAAKSRGRDIVCMDRVEEILEAAADHPLSRDNKWKVQKIETPLLICD 120
Db 133 IDGLRNIWIIKPAAKSRGRDIVCMDRVEEILEAAADHPLSRDNKWKVQKIETPLLICD 192

Qy 121 TKFDIRQWFLVTDWNPPLTIWFYKESYLRFSTQRFSLDKLDSAIHLCNNAVQKYLKNDVGR 180
Db 193 TKFDIRQWFLVTDWNPPLTIWFYKESYLRFSTQRFSLDKLDSAIHLCNNAVQKYLKNDVGR 252

Qy 181 SP LLPAHNMWTSRFOEYLQROGRGAVMGSVIYPSMKKAIAHAMKVAQDHVEPRKNSFEL 240
Db 253 SP LLPAHNMWTSRFOEYLQROGRGAVMGSVIYPSMKKAIAHAMKVAQDHVEPRKNSFEL 312

Qy 241 YGADFVLGRDPRFWLIEINSSPTMHPSTPVTQAQLCAQVQEDTIKVAVD RSCDI 293
Db 313 YGADFVLGRDPRFWLIEINSSPTMHPSTPVTQAQLCAQVQEDTIKVAVD RSCDI 365
```

```
RESULT 5
US-10-635-977-2
; Sequence 2, Application US/10635977
; Publication No. US20040171131A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635,977
```

; CURRENT FILING DATE: 2003-08-07
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S.10/615,659
; PRIOR FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 541
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-635-977-2

Query Match 100.0%; Score 1560; DB 4; Length 541;
Best Local Similarity 100.0%; Pred. No. 6.3e-151;
Matches 293; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 EDIDTSADAVEDLTEAEWEDLTQQYSLVHGDAFISNRNYSQOALLNRITSVNPQTD 60
DB 73 EDIDTSADAVEDLTEAEWEDLTQQYSLVHGDAFISNRNYSQOALLNRITSVNPQTD 132
QY 61 IDGLRNIIWIKPAAKSRGRDIVCMRDVEEILELAADHPLSRDNKWWVQKYEITPLLICD 120
DB 133 IDGLRNIIWIKPAAKSRGRDIVCMRDVEEILELAADHPLSRDNKWWVQKYEITPLLICD 192
QY 121 TKFDIRQWFLVTDWNPPLTIWFYKESYLRSTQRFSLDKLDSAIHLCLNNAVQKYLKNDVGR 180
DB 193 TKFDIRQWFLVTDWNPPLTIWFYKESYLRSTQRFSLDKLDSAIHLCLNNAVQKYLKNDVGR 252
QY 181 SPLLPAHNNMTSTRFOEYLQGRGAVGWSVIYPSMKKAIHAAMKVAQDHVPRKNSFEL 240
DB 253 SPLLPAHNNMTSTRFOEYLQGRGAVGWSVIYPSMKKAIHAAMKVAQDHVPRKNSFEL 312
QY 241 YGADFVLGRDRFRPWLIEINSSPTMHPSTPTVTAQLCAQVOEDTIKVAVDNRSCDI 293
DB 313 YGADFVLGRDRFRPWLIEINSSPTMHPSTPTVTAQLCAQVOEDTIKVAVDNRSCDI 365

RESULT 6
US-10-635-977-13
; Sequence 13, Application US/10635977
; Publication No. US20040171131A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635,977
; PRIOR FILING DATE: 2003-08-07
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S.10/615,659
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 13
; LENGTH: 541
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-635-977-13

Query Match 100.0%; Score 1560; DB 4; Length 541;
Best Local Similarity 100.0%; Pred. No. 6.3e-151;
Matches 293; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 EDIDTSADAVEDLTEAEWEDLTQQYSLVHGDAFISNRNYSQOALLNRITSVNPQTD 60
DB 73 EDIDTSADAVEDLTEAEWEDLTQQYSLVHGDAFISNRNYSQOALLNRITSVNPQTD 132
QY 61 IDGLRNIIWIKPAAKSRGRDIVCMRDVEEILELAADHPLSRDNKWWVQKYEITPLLICD 120
DB 133 IDGLRNIIWIKPAAKSRGRDIVCMRDVEEILELAADHPLSRDNKWWVQKYEITPLLICD 192

QY 121 TKFDIRQWFLVTDWNPPLTIWFYKESYLRSTQRFSLDKLDSAIHLCLNNAVQKYLKNDVGR 180
DB 193 TKFDIRQWFLVTDWNPPLTIWFYKESYLRSTQRFSLDKLDSAIHLCLNNAVQKYLKNDVGR 252
QY 181 SPLLPAHNNMTSTRFOEYLQGRGAVGWSVIYPSMKKAIHAAMKVAQDHVPRKNSFEL 240
DB 253 SPLLPAHNNMTSTRFOEYLQGRGAVGWSVIYPSMKKAIHAAMKVAQDHVPRKNSFEL 312
QY 241 YGADFVLGRDRFRPWLIEINSSPTMHPSTPTVTAQLCAQVOEDTIKVAVDNRSCDI 293
DB 313 YGADFVLGRDRFRPWLIEINSSPTMHPSTPTVTAQLCAQVOEDTIKVAVDNRSCDI 365

RESULT 7
US-10-615-659-4
; Sequence 4, Application US/10615659
; Publication No. US20040157234A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; PRIOR FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4
; LENGTH: 242
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-615-659-4

Query Match 79.9%; Score 1247; DB 4; Length 242;
Best Local Similarity 100.0%; Pred. No. 2.8e-119;
Matches 233; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 61 IDGLRNIIWIKPAAKSRGRDIVCMRDVEEILELAADHPLSRDNKWWVQKYEITPLLICD 120
DB 1 IDGLRNIIWIKPAAKSRGRDIVCMRDVEEILELAADHPLSRDNKWWVQKYEITPLLICD 60
QY 121 TKFDIRQWFLVTDWNPPLTIWFYKESYLRSTQRFSLDKLDSAIHLCLNNAVQKYLKNDVGR 180
DB 61 TKFDIRQWFLVTDWNPPLTIWFYKESYLRSTQRFSLDKLDSAIHLCLNNAVQKYLKNDVGR 120
QY 181 SPLLPAHNNMTSTRFOEYLQGRGAVGWSVIYPSMKKAIHAAMKVAQDHVPRKNSFEL 240
DB 121 SPLLPAHNNMTSTRFOEYLQGRGAVGWSVIYPSMKKAIHAAMKVAQDHVPRKNSFEL 180
QY 241 YGADFVLGRDRFRPWLIEINSSPTMHPSTPTVTAQLCAQVOEDTIKVAVDNRSCDI 293
DB 181 YGADFVLGRDRFRPWLIEINSSPTMHPSTPTVTAQLCAQVOEDTIKVAVDNRSCDI 233

RESULT 8
US-10-635-977-4
; Sequence 4, Application US/10635977
; Publication No. US20040171131A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635,977
; PRIOR FILING DATE: 2003-08-07
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S.10/615,659
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4

Db 2 DIDKDLAPLYLTPEGWSLFLQRYVQVHGAELRHLDLTVQVQCEDILOQLQAVVPQIDM 61
QY 62 DGLRNIIWIKPAKSRGRDIVCMRVVEIELEAAADHPLSRDNKWKVQKIETPLLCDT 121
Db 62 EGDRIWIVKPGAKSRGRGIMCNDHLEMLKLVNGNPFVMDKGVQVQKIETPLLCDT 121
QY 122 KFDIRQWFLVTDWNPPLTIWFKYKESYLRFTQSFSLDKLDSAIHLGNNAVQKYLKNDVGRS 181
Db 122 KFDLRQWFLVTDWNPPLTIWFKYKESYLRFTQSFSLDKLDSAIHLGNNSIQKHLNSCHRH 181
QY 182 PLLPAHNNMTSTRFOEYLQROGRGAVGWSVIYPSMKKAIAHAMKVAQDHVPRKNSFELY 241
Db 182 PLLPDPNNWSSORFOAHLQEMGAPNAWSTIIVPGMKDAVIHALQTSQDVTQCRKASFELY 241
QY 242 GADFVLRGRFRPWLIEINSSPTMHPSTPTVTAQLCAQVQEDTIKVAVDKSD 292
Db 242 GADFVFGEDFQFWLIEINASPTMAPSTAVTARLCAGVQADTLRVVIDRMLD 292

RESULT 12

US-10-275-595A-5

; Sequence 5, Application US/10275595A

; Publication No. US20040078804A1

; GENERAL INFORMATION:

; APPLICANT: YUE, Henry

; APPLICANT: TANG, Y. Tom

; APPLICANT: AU-YOUNG, Janice

; APPLICANT: LU, Dyung Aina R.

; APPLICANT: BAUGHN, Mariah R.

; APPLICANT: HILLMAN, Jennifer L.

; APPLICANT: AZIMZAI, Yalda

; APPLICANT: LAL, Preeti

; APPLICANT: YAO, Monique G.

; APPLICANT: BANDMAN, Olga

; APPLICANT: BURFORD, Neil

; APPLICANT: BATRA, Sajeev

; APPLICANT: KEARNEY, Liam

; APPLICANT: POLICKY, Jennifer L.

; TITLE OF INVENTION: CYTOSKELETON-ASSOCIATED PROTEINS

; FILE REFERENCE: PF-0772 USN

; CURRENT APPLICATION NUMBER: US/10/275,595A

; CURRENT FILING DATE: 2003-06-13

; PRIOR APPLICATION NUMBER: US 60/201,960

; PRIOR FILING DATE: 2000-05-05

; PRIOR APPLICATION NUMBER: US 60/202,729

; PRIOR FILING DATE: 2000-05-08

; PRIOR APPLICATION NUMBER: US 60/209,705

; PRIOR FILING DATE: 2000-06-05

; PRIOR APPLICATION NUMBER: US 60/210,149

; PRIOR FILING DATE: 2000-06-07

; PRIOR APPLICATION NUMBER: US 60/213,215

; PRIOR FILING DATE: 2000-06-21

; NUMBER OF SEQ ID NOS: 68

; SOFTWARE: PERL Program

; SEQ ID NO 5

; LENGTH: 488

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc_feature

; OTHER INFORMATION: Incyte ID No: 2156553CD1

US-10-275-595A-5

Query Match 58.2%; Score 908; DB 4; Length 488;
Best Local Similarity 56.7%; Pred. No. 4.9e-84;
Matches 165; Conservative 48; Mismatches 78; Indels 0; Gaps 0;
QY 2 DIDTSADAVEDLTEAEWEDLTQQYSLVHGDAFISNSRNFYSQCALNRIITSVNPQIDI 61
Db 2 DIDKDLAPLYLTPEGWSLFLQRYVQVHGAELRHLDLTVQVQCEDILOQLQAVVPQIDM 61
QY 62 DGLRNIIWIKPAKSRGRDIVCMRVVEIELEAAADHPLSRDNKWKVQKIETPLLCDT 121

; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 617
; LENGTH: 362
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (307)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-298-617

Query Match 57.5%; Score 897; DB 3; Length 362;
Best Local Similarity 57.5%; Pred. No. 4.4e-83;
Matches 161; Conservative 47; Mismatches 72; Indels 0; Gaps 0;

QY	13	LTEAEWEDLTQQYYSLVHGDAFISNRYFSQCALNLRITSVNPQTDIDGLNIIWIKP	72
DB	30	LTPGWSLFLQRYQVHVGEAELRHLDTVQRCEDILQQQAVVPQIDMEGDRNIWIKP	89
QY	73	AAKSRGDIVCMRDRVEILEAAADHPLSRDKNVVKQYIETPLLICDTKFDIRQWFLVT	132
DB	90	GAKSRGIMCMHLEEMKLVNGNPVVMKDGKVVQKYIERPLLIFFGTKFDLRQWFLVT	149
QY	133	DWNPLTIWFYKESYLRFSSTORFSLDKLDSAIHLCNNAVOKYLNVDVGRSPLLPAHNMWTS	192
DB	150	DWNPLTVWFYRDSYIRFSTQPFSLKNLDSNVHLCNNSIQKHLENSCHRHPLLPDPNNWSS	209
QY	193	TRFOEVLQROGRGAVGWSVIYPSMKKAIHAHAKVAQDHVEPRKNSFELYGADFVLGRDFR	252
DB	210	QRFOAHLQENGAPNASTIIVPGMKDAVIHALQTSQDTVQCRKASFELYGADFVFGEDFQ	269
QY	253	PWLIENSPTMHPSTPVTQAQLCAQVOEDTIKVAVDNRSCD	292
DB	270	PWLIENASPTWAPSTAVTARLCAGVQADTLRVVIDRXLD	309

RESULT 15

US-10-102-806-617
; Sequence 617, Application US/10102806
; Publication No. US20030054421A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA103P1C1
; CURRENT APPLICATION NUMBER: US/10/102.806
; CURRENT FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: 09/925,298
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05881
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 846
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 617
; LENGTH: 362
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (307)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-102-806-617

Query Match 57.5%; Score 897; DB 4; Length 362;
Best Local Similarity 57.5%; Pred. No. 4.4e-83;
Matches 161; Conservative 47; Mismatches 72; Indels 0; Gaps 0;

QY	13	LTEAEWEDLTQQYYSLVHGDAFISNRYFSQCALNLRITSVNPQTDIDGLNIIWIKP	72
DB	30	LTPGWSLFLQRYQVHVGEAELRHLDTVQRCEDILQQQAVVPQIDMEGDRNIWIKP	89
QY	73	AAKSRGDIVCMRDRVEILEAAADHPLSRDKNVVKQYIETPLLICDTKFDIRQWFLVT	132

DB	90	GAKSRGIMCMHLEEMKLVNGNPVVMKDGKVVQKYIERPLLIFFGTKFDLRQWFLVT	149
QY	133	DWNPLTIWFYKESYLRFSSTORFSLDKLDSAIHLCNNAVOKYLNVDVGRSPLLPAHNMWTS	192
DB	150	DWNPLTVWFYRDSYIRFSTQPFSLKNLDSNVHLCNNSIQKHLENSCHRHPLLPDPNNWSS	209
QY	193	TRFOEVLQROGRGAVGWSVIYPSMKKAIHAHAKVAQDHVEPRKNSFELYGADFVLGRDFR	252
DB	210	QRFOAHLQENGAPNASTIIVPGMKDAVIHALQTSQDTVQCRKASFELYGADFVFGEDFQ	269
QY	253	PWLIENSPTMHPSTPVTQAQLCAQVOEDTIKVAVDNRSCD	292
DB	270	PWLIENASPTWAPSTAVTARLCAGVQADTLRVVIDRXLD	309

Search completed: April 4, 2006, 12:58:00
Job time : 43.2072 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2006, 12:46:06 ; Search time 11.4368 Seconds
(without alignments)
2118.072 Million cell updates/sec

Title: US-10-635-977-2_COPY_73_365

Perfect score:

Sequence: 1 EDITSADAVEDLTEAEWED.....LCAQVQEDTIKVAVDRSCDI 293

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 s

```
Database : Issued Patents AA:*
1: /cgn2_6/ptodata/1/iaa/5_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/6_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/H_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/PTCUS_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/RE_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query			ID	Description	
	Score	Match	Length			
1	271.5	17.4	592	2	US-10-104-047-3371	Sequence 3371, Ap
2	256.5	16.4	439	2	US-10-104-047-3658	Sequence 3658, Ap
3	212	13.6	561	2	US-09-270-767-46703	Sequence 46703, A
4	180	11.5	404	2	US-10-104-047-3621	Sequence 3621, A
5	154.5	9.9	260	2	US-09-270-767-62411	Sequence 62411, Ap
6	154.5	9.9	507	2	US-09-270-767-46787	Sequence 46787, A
7	151	9.7	259	2	US-09-248-796A-14603	Sequence 14603, A
8	144.5	9.3	753	2	US-09-538-092-31	Sequence 31, Appl
9	126	8.1	305	2	US-09-270-767-46577	Sequence 46577, A
10	102	6.5	55	2	US-09-270-767-62317	Sequence 62317, A
11	100.5	6.4	398	2	US-09-902-540-12305	Sequence 12305, A
12	94	6.0	626	1	US-08-596-300A-7	Sequence 7, Appl
13	94	6.0	626	1	US-08-596-300A-14	Sequence 14, Appl
14	93.5	6.0	816	2	US-09-248-796A-20939	Sequence 20939, A
15	90.5	5.8	732	2	US-08-671-757A-7	Sequence 7, Appl
16	90.5	5.8	732	2	US-08-671-757A-8	Sequence 8, Appl
17	90.5	5.8	732	2	US-09-015-078-7	Sequence 7, Appl
18	90.5	5.8	732	2	US-09-015-078-8	Sequence 8, Appl
19	90.5	5.8	732	2	US-10-238-977A-7	Sequence 7, Appl
20	90.5	5.8	732	2	US-10-238-977A-8	Sequence 8, Appl
21	87.5	5.6	607	2	US-09-489-039A-12140	Sequence 12140, A
22	87	5.6	218	2	US-09-328-352-6389	Sequence 6389, Ap
23	87	5.6	884	6	5208144-8	Patent No. 5208144
24	85	5.4	938	2	US-09-489-039A-13504	Sequence 13504, A
25	83.5	5.4	268	1	US-07-857-224B-33	Sequence 33, Appl
26	83	5.3	418	2	US-08-908-436-6	Sequence 6, Appl
27	83	5.3	418	2	US-09-291-289-4	Sequence 4, Appl

RESULT 1

US-10-104-047-3371
US-10-104-047-3371 ; Sequence 3371, Application US/10104047

Patent No. 6943241
GENERAL INFORMATION:

; APPLICANT: HELIX RESEARCH INSTITUTE
 ; TITLE OF INVENTION: No. 6943241e1 full length cdna

FILE REFERENCE: H1-A0105

; CURRENT APPLICATION NUMBER: US/100

;
CURRENT FILING DATE: 2005
PRIORITY DATE: 2005

;
PRIOR APPLICATION
PRIOR PENDING PATE

; PRIOR FILING DATE:
 ; NUMBER OF SEC ID NOS. 1000

```

; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatonIn Ver 3.1

```

; SOFTWARE: PAD
: CEO ID NO 3371

```

; SEQ ID NO 337
:
: LENGTH: 592

```

```

; LENGTH: 3
; TYPE: PRT

```

TYPE: PRI
ORGANISM: Homo sapiens

US-10-104-047-3371

Query Match	17.4%	Score 271.5;	DB 2;	Length 592;
Best Local Similarity	29.8%	Pred. No. 2.9e-21;		
Matches 72;	Conservative 43;	Mismatches 80;	Indels 47;	Gaps 8;

QY 41 YFSQCOALLNRITSVNPQTDIDGLRNIWI KPAAKSRGRDIVCM DRVVEIILELAAADHPL 100

[illegible]

Db 195 YFQERQMLGTK-----HSYWICKPAELSRGRGILIFSDFKDFI----- 232

.....

101 SRDNKVVQKYIETPLLCIDTKFDIRQWFLVTDWNPLTIWFYKESYLRFSTQRFSLDKL- 159 QY

[illegible]

233 -FDDMYIVQKYISNPLLI GRYKCDLRIYVCVTGFKPLTIYVYQEGLVRFATEKFDLSNLQ 291

Qy 160 DSAIHL

\vdots

RESULT 2

US-10-104-047-3658
US-10-104-047-3658, Application US/10104047

Patent No. 6943241
GENERAL INFORMATION:

```
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. 6943241el full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3658
; LENGTH: 439
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-104-047-3658

Query Match 16.4%; Score 256.5; DB 2; Length 439;
Best Local Similarity 25.5%; Pred. No. 8.9e-20;
Matches 84; Conservative 56; Mismatches 109; Indels 81; Gaps 13;

QY 10 VEDLTAEW-----DLT-----QYYSLVHGDAFISNSRYFSQC 45
DB 65 VEYKDEGEWDFYWCDSWLENDFDYMDEHVRIASHFRNHYTELRKNYMWKNLKRKQL 124
QY 46 QALLNRTSVN-----POT-----IDGLRN-----IWIIPAAKSRGRDIVCMRDVE 88
DB 125 ERAGKLEAAKCDFFPKTFEMPCEYHLFVEFRKNPGCITWIMKPVARSQKGIFLFRLLK 184
QY 89 EILELAAADHPLSRDNK-----WVQYIETPLLICDTKFDIROWFLVTDWNPITWIF 141
DB 185 DIVDM-RKOTRSSDDQKDDIPVENYVAQRYIENPYLLIGRKFDRVVVLVMS-----V 236
QY 142 YKSYLRFSTQFSLDKLSAHLCCNNAVQKYLKNDVGRSPLPAHN-----WMTSTRFQ 196
DB 237 FAECLLWSGHR-----QDVHLTNVAVOK-----TSPDYHPKKGCKWTKLQFR 279
QY 197 EYL-ORCGAVGMSVTPSMKKAIAHAMKVAQDHVEPRKNSFELYGADFVLGRDFPWL 255
DB 280 QYLASKGPEAV--ETLFRIDINIFVKSLSQSVQKVIISDKHCFELYGIDYILIDQDLKPWL 337
QY 256 IEINSSPTMHPSTPVTQAQLCAQVQEDTIKV 285
DB 338 LEVNASPLTASSQEDYELKTCLEDTLHV 367

RESULT 3
US-09-270-767-46703
; Sequence 46703, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 46703
; LENGTH: 561
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-46703

Query Match 13.6%; Score 212; DB 2; Length 561;
Best Local Similarity 29.2%; Pred. No. 1.3e-14;
Matches 57; Conservative 37; Mismatches 75; Indels 26; Gaps 6;

QY 107 WVQYIETPLLICDTKFDIROWFLVTDWNPITWIFKESYLRFSTQFSLDKL 163
DB 1 VWSKYIVDPLCIDGHKCDLRVYLVTSFDPYIYLYEGIVRLATVYKIDRHADLNWPCM 60
QY 164 HLCNNAVQKYLKN-----DVGRSPLPAHNMTSTRFQYLRQGG---RGAVWG- 209
DB 61 HLCNYSINKYHSDYIRSSDAQDEDVG-----HKWLSALLRHLKLSQCDTRQLMNI 112

; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. 6943241el full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3621
; LENGTH: 404
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-104-047-3621

Query Match 11.5%; Score 180; DB 2; Length 404;
Best Local Similarity 25.6%; Pred. No. 2.9e-11;
Matches 63; Conservative 33; Mismatches 84; Indels 66; Gaps 12;

QY 4 DTSADAVEDTEAEWEDLTQQYYSLVHGDAFISNSRYFSQCALNRTISVNPOT---D 60
DB 183 DAAAPALEDI---PW-----TSPGYLRPQVRL--RMEFFPETYRLD 219
QY 61 IDGLR-----NIWIIPAAKSRGRDIVCMRDVEIILELAA-----ADHPLSRDNK 105
DB 220 LKHERAEFTLFDDETQIWKCTASNOGKGFILRNQEEVAAQAQTRSMEDDPIHHTP 279
QY 106 W-----WVQYIETPLLICDTKFDIROWFLVTDWNPITWIFKESYLRFSTQFSLDKL 159
DB 280 FRGQPARVQRYIQNPLLVDRKFDVRSYLLIACTTPTMI-FFGHGVARLTLSLYDPHSS 338
QY 160 DSAIHLCCNNAVQKYLKNDVGRSP---LLPAHNMTSTRFQEYL-----QRQGVAVGVS 211
DB 339 DLGGHLTNQFMQK-----KSPLYMLLKEHTVMSMEHLNRYISDTFWKARGLAKDW--- 388
QY 212 IYFSMK 217
DB 389 VFTTLK 394

RESULT 5
US-09-270-767-62411
; Sequence 62411, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 62411
; LENGTH: 260
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-62411
```


INFORMATION FOR SEQ ID NO: 7;
SEQUENCE CHARACTERISTICS:
LENGTH: 626 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
FRAGMENT TYPE:
US-08-596-300A-7

Query Match 6.0%; Score 94; DB 1; Length 626;
Best Local Similarity 20.1%; Pred. No. 0.24;
Matches 72; Conservative 40; Mismatches 126; Indels 120; Gaps 17;
QY 4 DTSADAVEDLTAEWEDLTQQYYSLVHGDFAFISNRNYSQCOA----- 47
DB 207 NTGFDLWEENVSSPFTTAAASHRALVEGSAFAKSVGSSCSACDAPILCFQQSFWSNS 266
QY 48 --LLNRITSVNPDIDG--LRNIWIKPAAKSRGRDI-----VCMRVEBIELEAAADHPL 100
DB 267 GYIISNFVNRSGKDINSVLTSIHNFDPAA--GCDVNTFQPCSDR-----ALANHKV 316
QY 101 SRDNK--W-----VVQKIETPLLICDTKFDIRQWFLVT-----DWNPLTIW 140
DB 317 VVDSMRFGVNSGRTAGKAAAVGRYAE-----DVYINGNPWYLATLAAAEQLYDAVYVW 370
QY 141 -----FYKESYLRFSTQRFSLDKLDSAIHLCCNNAVOKY----- 173
DB 371 KKQGSITVTSTSLAFKDLVPSVSTGYTSS--SSSTYTAIINAVTYADGFVDIIVAQYTP 428
QY 174 -----LKNVGRSPLLPAAHNMWTSRFBQYLRQGRGAV---WGSVIYPSMKKAIAH 222
DB 429 SDGSLAEQFDKDSG--APLSATHLTWSYASFLSAAARRA--GIVPPSWGAAASNSLPGSCS- 485
QY 223 AMKVAQDHVEPRKNSFELYGADFLGRDPRFWLIEINSPTWHPSTPTVTAQLCAQVQE 280
DB 486 ASTVAGSYATATATSF-----PANLTPASTTTPPTQTGCAADHE 525

RESULT 13
US-08-596-300A-14
Sequence 14, Application US/08596300A
Patent No. 5834191
GENERAL INFORMATION:
APPLICANT: Keith Weinstein et al
TITLE OF INVENTION: Production of Heterologous Peptides
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/596.300A
FILING DATE: 13-FEB-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 1321-1-001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:
LENGTH: 626 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
FRAGMENT TYPE:
US-08-596-300A-14

Query Match 6.0%; Score 94; DB 1; Length 626;
Best Local Similarity 20.1%; Pred. No. 0.24;
Matches 72; Conservative 40; Mismatches 126; Indels 120; Gaps 17;
QY 4 DTSADAVEDLTAEWEDLTQQYYSLVHGDFAFISNRNYSQCOA----- 47
DB 207 NTGFDLWEENVSSPFTTAAASHRALVEGSAFAKSVGSSCSACDAPILCFQQSFWSNS 266
QY 48 --LLNRITSVNPDIDG--LRNIWIKPAAKSRGRDI-----VCMRVEBIELEAAADHPL 100
DB 267 GYIISNFVNRSGKDINSVLTSIHNFDPAA--GCDVNTFQPCSDR-----ALANHKV 316
QY 101 SRDNK--W-----VVQKIETPLLICDTKFDIRQWFLVT-----DWNPLTIW 140
DB 317 VVDSMRFGVNSGRTAGKAAAVGRYAE-----DVYINGNPWYLATLAAAEQLYDAVYVW 370
QY 141 -----FYKESYLRFSTQRFSLDKLDSAIHLCCNNAVOKY----- 173
DB 371 KKQGSITVTSTSLAFKDLVPSVSTGYTSS--SSSTYTAIINAVTYADGFVDIIVAQYTP 428
QY 174 -----LKNVGRSPLLPAAHNMWTSRFBQYLRQGRGAV---WGSVIYPSMKKAIAH 222
DB 429 SDGSLAEQFDKDSG--APLSATHLTWSYASFLSAAARRA--GIVPPSWGAAASNSLPGSCS- 485
QY 223 AMKVAQDHVEPRKNSFELYGADFLGRDPRFWLIEINSPTWHPSTPTVTAQLCAQVQE 280
DB 486 ASTVAGSYATATATSF-----PANLTPASTTTPPTQTGCAADHE 525

RESULT 14
US-09-248-796A-20939
Sequence 20939, Application US/09248796A
Patent No. 6747137
GENERAL INFORMATION:
APPLICANT: Keith Weinstein et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBIC
TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.132
CURRENT APPLICATION NUMBER: US/09/248.796A
CURRENT FILING DATE: 1999-02-12
PRIOR APPLICATION NUMBER: US 60/074,725
PRIOR FILING DATE: 1998-02-13
PRIOR APPLICATION NUMBER: US 60/096,409
PRIOR FILING DATE: 1998-08-13
NUMBER OF SEQ ID NOS: 28208
SEQ ID NO 20939
LENGTH: 816
TYPE: PRT
ORGANISM: Candida albicans
US-09-248-796A-20939

Query Match 6.0%; Score 93.5; DB 2; Length 816;
Best Local Similarity 20.8%; Pred. No. 0.42;
Matches 64; Conservative 45; Mismatches 125; Indels 73; Gaps 13;
QY 8 DAVEDLTAEWEDLTQQYYSLVHGD-----AFISNRNYSF----- 42
DB 348 DAFQDLNLSNTGDVV---IKASDGFVAHAPVLKARSAPFETLLSERWDTBEKGNVQYVD 404
QY 43 -----SQCCALLNRITSVNPDID-----GLRNWIKPAAKSRGRDIVCMRVEI 90
DB 405 FTGLTKFQVTLIRHLRYGVNSLSDCFQYDFGSKDYPFINDLLIELIEVADELLFQLKSV 464

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2006, 12:43:06 ; Search time 44.3857 Seconds
(without alignments)
4657.352 Million cell updates/sec

Title: US-10-635-977-2_COPY_73_365
Perfect score: 1560
Sequence: 1 EDITSDAVALDELTAEWED.....LCAQVQEDTIKVAVDKSDI 293

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues
Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt 05.80:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match	Length	DB	ID	Description
1	1447	92.8	747	2	Q5JZ84	HUMAN	Q5JZ84 homo sapien
2	1303	83.5	518	2	Q8C0V2	MOUSE	Q8C0V2 mus musculus
3	1303	83.5	781	2	Q8CON7	MOUSE	Q8CON7 mus musculus
4	908	58.2	352	1	TTLI3	HUMAN	Q9V4I7 homo sapien
5	908	58.2	352	2	Q6AWA3	HUMAN	Q6AWA3 homo sapien
6	908	58.2	352	2	Q4KWS8	HUMAN	Q4KWS8 homo sapien
7	908	58.2	434	2	Q8NDN8	HUMAN	Q8NDN8 homo sapien
8	908	58.2	744	2	Q9H876	HUMAN	Q9H876 homo sapien
9	893	57.2	704	2	Q8BV51	MOUSE	Q8BV51 mus musculus
10	855.5	54.8	534	2	Q4RY08	TETNG	Q4RY08 tetraodon n
11	805	51.6	261	2	Q58CT2	BOVIN	Q58CT2 bos taurus
12	805	51.6	266	2	Q922T0	MOUSE	Q922T0 mus musculus
13	673	43.1	572	2	Q7Q156	ANOGA	Q7Q156 anopheles g
14	636	40.8	992	2	Q9VM91	DROME	Q9VM91 drosophila
15	539.5	34.6	501	2	Q7PMD3	ANOGA	Q7PMD3 anopheles g
16	515.5	33.0	719	2	Q9VM92	DROME	Q9VM92 drosophila
17	515.5	33.0	756	2	Q5BHY1	DROME	Q5BHY1 drosophila
18	439	28.1	281	2	Q6ZU95	HUMAN	Q6ZU95 homo sapien
19	413	26.5	331	2	Q5TNZ9	ANOGA	Q5TNZ9 anopheles g
20	385	24.7	101	2	Q9GGG8	HUMAN	Q9GGG8 homo sapien
21	344	22.1	523	2	Q54TUI	DICDI	Q54TUI dictyosteli
22	331	21.2	1075	2	Q7QTI3	GIALA	Q7QTI3 giardia lam
23	330	21.2	794	2	Q6BFH6	PARTI	Q6BFH6 paramacium
24	327.5	21.0	461	2	Q641W7	RAT	Q641W7 rattus norv
25	325	20.8	464	2	Q9D570	MOUSE	Q9D570 mus musculus
26	324	20.8	496	2	Q961I9	DROME	Q961I9 drosophila
27	324	20.8	496	2	Q9VX74	DROME	Q9VX74 drosophila
28	322	20.6	375	2	Q4V8C1	RAT	Q4V8C1 rattus norv
29	306.5	19.6	403	2	Q7QZC8	GIALA	Q7QZC8 giardia lam
30	299	19.2	1339	2	Q5F498	CHICK	Q5F498 gallus gall
31	297.5	19.1	827	2	Q9VKL9	DROME	Q9VKL9 drosophila

ALIGNMENTS

RESULT 1

Q5JZ84 HUMAN PRELIMINARY; PRT; 747 AA.
AC Q5JZ84
DT 10-MAY-2005 (Tremblrel. 30, Created)
DT 10-MAY-2005 (Tremblrel. 30, Last sequence update)
DE OTTHUMP0000028514 (Fragment).
GN ORFNames=RP3-355C18.2-002;
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
RX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Cobley V.;
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL022327; CA142686.1; -; Genomic DNA.
DR GO; GO:0004835; F:tubulin-tyrosine ligase activity; IEA.
DR GO; GO:0006464; P:protein modification; IEA.
DR InterPro; IPR004344; Tub_tyr_lygase.
DR Pfam; PF03133; TTL; 1.
FT NON_TER 1
SQ SEQUENCE 747 AA; 83854 MW; F90948E159BBE589 CRC64;

Query Match 92.8%; Score 1447; DB 2; Length 747;
Best Local Similarity 89.2%; Pred. No. 1e-114; 6; Indels 24; Gaps 4;
Matches 281; Conservative 4; Mismatches 6; Indels 24; Gaps 4;

QY 1 EDITSDAVALDELTAEWEDLTQYYISLVH---GDAFI-----SNSRNYF 42
Db 259 EDITSDAVALDELTAEWEDLTQYYISLVQVPLGSSIVLCIFKIQKVMSPFPPTARD--- 316
QY 43 SQCCALLNRITSVNPQTDIGLRNIWIIPKAASRGR---DIVCMRVVEEILELAADH 98
Db 317 RQCQALLNRITSVNPQTDIGLRNIWIIPKAASRGRGSPDIVCMRVVEEILELAADH 376
QY 99 PLSRDNKVVQKYLETPLLICDTKFDIQWFLVTDNPLTIWFYKESYLRFSTQFSLDK 158
Db 377 PLSRDNKVVQKYLETPLLICDTKFDIQWFLVTDNPLTIWFYKESYLRFSTQFSLDK 436
QY 159 LDSAHLCCNNAVQKYLKNDVGRSPLLPAHNNMTSTRFOYLQQRQGRGAVGWSVIYPSMKK 218
Db 437 LDSAHLCCNNAVQKYLKNDVGRSPLLPAHNNMTSTRFOYLQQRQGRGAVGWSVIYPSMKK 496
QY 219 AIAHAMKVAQDHVSPRKNSEFELGYADFLVGRDFRPWLIEINSSPTMHPSTPVTQAQCV 278
Db 497 AIAHAMKVAQDHVSPRKNSEFELGYADFLVGRDFRPWLIEINSSPTMHPSTPVTQAQCV 556
QY 279 QEDTIKVAVDKSDI 293
Db 557 QEDTIKVAVDKSDI 571

Q8IGW4 drosophila
Q8IPB2 drosophila
Q8C125 mus musculus
Q8CFV5 mus musculus
Q8CHB8 mus musculus
Q8PLV5 homo sapien
Q9UPZ4 homo sapien
Q6EMB2 homo sapien
Q6EEF3 cercopithec
Q5R978 pongo pygma
Q5VX47 homo sapien
Q8T417 drosophila
Q9VQV6 drosophila

```

RESULT 2
Q8COV2 MOUSE
ID Q8COV2 MOUSE PRELIMINARY; PRT; 518 AA.
AC Q8COV2;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE Mus musculus adult male testis cDNA, RIKEN full-length enriched
DE library, clone:4930524K07 product:hypothetical Tubulin-tyrosine ligase
DE containing protein, full insert sequence.
GN Name:1700019P01Rik;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]_TaxID=10090;
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RC MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RA "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
[2]
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RC MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Waghio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,
RA Hayashizaki Y.;
RA "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
[3]
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RC MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
[5]
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RC MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoaka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,

```

```

RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
[6]
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Kato H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK029745; BAC26595.1; -; mRNA.
DR MGI; MGI:1922902; 170019P01Rik.
DR GO; GO:0016874; P.ligase activity; IEA.
DR GO; GO:0004835; F.tubulin-tyrosine ligase activity; IEA.
DR GO; GO:0006464; P.protein modification; IEA.
DR InterPro; IPR004344; Tub_tyf_ligase.
DR Pfam; PF03133; TTL; 1.
DR Hypothetical protein; Ligase.
KW SEQUENCE 518 AA; 59319 MW; 540C045659FAC0D6 CRC64;
SQ
Query Match 83.5%; Score 1303; DB 2; Length 518;
Best Local Similarity 80.5%; Pred. No. 1.3e-102;
Matches 239; Conservative 29; Mismatches 25; Indels 4; Gaps 1;
QY 1 EDIDTSADAVEDITAEAWEDLTQOYSLVHGDAFINSRNYFQCCALLNRITSVNPQTD 60
DB 129 EDIDVSEASTEALSEEWNDLTQOYLLVHGNAISITDSKSYFAQCQALLSKISSVPQTE 188
QY 61 IDGLRNIIWIKPAKSGRDIVCMRVEIFILEALAAADHPLSRDNKWKVQKIETPLIICD 120
DB 189 IDGIRNIWIKPAKSGRDIVCMRVEIILSLVAADSQTTKONKWKVQKIETPLIYD 248
QY 121 TKFDIQWFLVTDWNPDLTTFYKESYLRFSTQFSLDKLSDAHLGNNAVQKYLKNDVGR 180
DB 249 TKFDIRQWFLVTDWNPDLTTFYKESYLRFSTQFSLDKLSDAHLGNNSIQRLKNDKER 308
QY 181 SPLLPAHNMWTSRPFQYLRQGRGAVWGSVIYPSMKKALAHAMKVAQDHVEPRKNSFEL 240
DB 309 SPLLPCHNMWTSRPFQYLRQGRGGTWSGIYPSMKRAVTNMRVAQDHVEARKNSFEL 368
QY 241 YGADFVLGRDPRFWLIEINSSPTMHPSTPVTALCAQVQSDTIKVAV-----DRSCDI 293
DB 369 YGADFILGRDFKFWLIEINSSPTMHPSTPVTALCAQVQSDTIKVAVVVDVDRKLDNRCDI 425
RESULT 3
Q8CON7 MOUSE
ID Q8CON7 MOUSE PRELIMINARY; PRT; 781 AA.
AC Q8CON7;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE Mus musculus adult male testis cDNA, RIKEN full-length enriched
DE library, clone:4933401B17 product:hypothetical Tubulin-tyrosine ligase
DE containing protein, full insert sequence.
GN Name:1700019P01Rik;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muroidae; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;

```

RN NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Testis;
 RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
 RA Carninci P., Hayashizaki Y.;
 RT "High-efficiency full-length cDNA cloning.";
 RL Meth. Enzymol. 303:19-44(1999).
 RN NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Testis;
 RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavani T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 RN NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Testis;
 RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
 RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes.";
 RL Genome Res. 10:1617-1630(2000).
 RN NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Testis;
 RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
 RA Konno H., Akiyama J., Nishi K., Katsunai T., Tashiro H., Itoh M.,
 RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
 RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-format
 sequencing pipeline with 384 multicapillary sequencer.";
 RL Genome Res. 10:1757-1771(2000).
 RN NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Testis;
 RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
 RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
 RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
 RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kagawa T.,
 RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
 RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
 RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
 RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,

RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
 RA Tagawa A., Takahashi F., Takaku-Akashira S., Takeda Y., Tanaka T.,
 RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL, AK030151; BAC26811.1; -, mRNA.
 DR Ensembl; ENSMUSG0000022386; Mus musculus.
 DR MGI; MGI:1922902; 1700019P01Rik.
 DR GO; GO:0016874; F:ligase activity; IEA.
 DR GO; GO:0004835; F:tubulin-tyrosine ligase activity; IEA.
 DR GO; GO:0006464; P:protein modification; IEA.
 DR InterPro; IPR000276; GPCR Rhodopsin.
 DR Pfam; PF03133; TTL; 1.
 DR InterPro; IPR004344; Tub_Tyr_ligase.
 DR PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; UNKNOWN_1.
 KW Hypothetical protein; Ligase.
 SQ SEQUENCE 781 AA; 89398 MW; 9E98793C3351C3DE CRC64;
 Query Match 83.5%; Score 1303; DB 2; Length 781;
 Best Local Similarity 80.5%; Pred. No. 2.2e-102;
 Matches 239; Conservative 29; Mismatches 25; Indels 4; Gaps 1;
 QY 1 EDIDTSADAVDELTEAEWEDLTQOYYSLVHGDATFISNSNYSOCOALLNRITSVNPOTD 60
 Db ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 276 EDIDVSEASTSEALSEEWNLDLTQOYLLVHGNSITDSKSYFAQCOALLSKISSVNPQTE 335
 QY 61 IDGLENIWIIKPAKSRGRDIVCMRDVEEILEAAADHPLSRDNKVVQKVIETPLLICD 120
 Db ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 336 IDGIRNIWIIKPAKSRGRDIVCMRDVENILSLVAADSTQTTKDNKVVQKVIETPLMIIYD 395
 QY 121 TKFDIROWFLVTDNPLTIWFYKESYLRFSQTFSLDKLDSAIHLGNNAVQKLVNDVGR 180
 Db ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 396 TKFDIROWFLVTDNPLTIWFYKESYLRFSQTFSLDKLDSAIHLGNNSIQRLKNDKER 455
 QY 181 SPLLPANHWSTTRFQYLRQGRGAVGVSIVPSMKRAIAHAMKVAODHVPKNSFEL 240
 Db ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 456 SPLLPCHNHWSTTRFQYLRQGRGTGWSIIYPSMKRAVTNMRVAQDHVEARKNSFEL 515
 QY 241 YGADPVLGRDPRPMLIENISPTMHPSTPTVTAQLCAQVQEDTIKAVV----DRSCDI 293
 Db ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 516 YGADPILGRDFKPMLEINISPTMHPSTPTVTAQLCAQVQEDTIKVVVDKLDKNCDI 572
 RESULT 4
 TTL3 HUMAN
 ID TTL3 HUMAN STANDARD; PRT; 352 AA.
 AC Q9V4R7; OSUI99;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 13-SEP-2005 (Rel. 48, Last annotation update)
 DE Tubulin tyrosine ligase-like protein 3 (HOTT1).
 GN Name=TTL3; ORFNames=PRO0207;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
 OC Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
 RC TISSUE=Testis;
 RL The German cDNA consortium;
 RG Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA] OF 61-352.
 RC TISSUE=Fetal liver;
 RA Yu Y., Zhang C., Luo L., Ouyang S., Zhang S., Li W., Wu J., Zhou S.,
 RA Liu M., He F.;
 RT "Functional prediction of the coding sequences of 50 new genes deduced
 by analysis of cDNA clones from human fetal liver.";
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: Contains 1 TTL domain.
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -

RP	NUCLEOTIDE SEQUENCE.
RC	TISSUE-PCR rescued clones;
RC	NIH MGC Project.
RL	Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
DR	EMBL; AC038939; CAD38794.1; --; mRNA.
DR	EMBL; BC038298; AAH98298.1; --; mRNA.
DR	GO; GO:0004835; F:tubulin-cytosine ligase activity; IEA.
DR	GO; GO:0006454; P:protein modification; IEA.
DR	InterPro; IP0004344; Tub_tyr_ligase.
DR	Pfam; PF03133; TTL; 1.
KW	Hypothetical protein.
SQ	SEQUENCE 434 AA; 49433 MW; 9E79E6CA08651CA1 CRC64;
	Query Match 58.2%; Score 908; DB 2; Length 434;
	Best Local Similarity 56.7%; Pred. No. 6.2e-69;
	Matches 165; Conservative 48; Mismatches 78; Indels 0; Gaps 0;
Qy	2 DIDTSADAVEDLTAENBDLTQQYYSLVHGDAFISNRNRYFSQCQALLNRITTSVNPQTDI 61
Db	2 DIDKDLAPLYLTPEGWSLFLQRYVVVHVEGAELRHLDTPQRCEDILQOLQAVVPQIDM 61
Qy	62 DGLRNIIKPAKSRGRDIDVCMRDVEISELELAADHPLSRKNKVVQKYIETPLLICDT 121
Db	62 EGDNRNIWIKFGAKSRGRGIMCMDHLEMLKLVNGPVVMKDGKVVQKYIERPLLI 121
Qy	122 KFDIRQFLVTDWNPITITWFKESYLRFSTORFSLDKLDSATHLCNNVQKYLKNDVGRS 181
Db	122 KFDIRQFLVTDWNPITITWFKESYLRFSTORFSLDKLDSATHLCNNVQKYLKNDVGRS 181
Qy	182 PLLPAHNNMTSTREFQYLQROGRGAVMGVSIVPSMKKATAHAMKVAQDHVEPRKNSFELY 241
Db	182 PLLPPDNWSSQRFQAHLEQENAPNAWSTIIIVPGMKDAVIHALQTSQDTVQCKASFELY 241
Qy	242 GADPVLGRDFRPWLIEINSSPTMPSTPTVTAQLCAQVEDTIKVAVDKSD 292
Db	242 GADPVGDFGDFQPLWLEINASPTMAPSTAVTARLCAGVQADTLRVVVIDRLD 292
RESULT 8	
Q9H876_HUMAN	
ID	Q9H876 HUMAN PRELIMINARY; PRT; 744 AA.
AC	Q9H876;
DT	01-MAR-2001 (TrEMBLrel. 16, Created)
DT	01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE	Hypothetical protein FLJ13898.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Euthera; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC	Homo.
OX	NCBI_TaxID=9606;
RN	[1]
RP	NUCLEOTIDE SEQUENCE.
RC	TISSUE=Thyroid gland;
RC	PubMed=14702039; DOI=10.1038/ng1285;
RA	Ota T., Suzuki Y., Nishikawa T., Ohtsuki T., Sugiyama T., Irie R.,
RA	Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA	Sekine M., Ohbayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
RA	Yamanoto J.-I., Saito K., Kawai Y., Isono Y., Nakamura Y.,
RA	Nagahari K., Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M.,
RA	Shiratori A., Sudo H., Hosoiti T., Kaku Y., Kodaira H., Kondo H.,
RA	Sugawara M., Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E.,
RA	Omura Y., Abe K., Kamihara K., Katsuda N., Sato K., Tanikawa M.,
RA	Yamazaki M., Ninomiya K., Ishibashi T., Yamashita H., Murakawa K.,
RA	Fujimori K., Tanai H., Kimata M., Watanabe M., Hiraoka S., Chiba Y.,
RA	Ishida S., Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotta T.,
RA	Kusano J., Kanehori K., Takahashi-Fujii A., Hara H., Tanase T.-O.,
RA	Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K., Arita M.,
RA	Inoue N., Musashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,
RA	Yoshikawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,
RA	Moriya S., Momiyama H., Satoh N., Takami S., Teraashima Y., Suzuki O.,
RA	Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
RA	Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami R.,

[illegible]

STRAIN=C57BL/6J; TISSUE=Thymus;
MEDLINE=21055660; PubMed=11217851; DOI=10.1038/35055500;
Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
Saito T., Okazaki Y., Gojibori T., Bono H., Kasukawa T., Saito R.,
Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
Fleischmann W., Gaasterland T., Glass C., King B., Kochiwa H.,
Kuenl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Bargh G.,
Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
Nordone P., King B., Rongwald M., Rodriguez I., Sakamoto N.,
Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohseki S.,
Hayashizaki Y.;
"Functional annotation of a full-length mouse cDNA collection.";
Nature 403:685-690(2001).
[3]
NUCLEOTIDE SEQUENCE.
STRAIN=C57BL/6J; TISSUE=Thymus;
The FANTOM Consortium.
"Analysis of the mouse transcriptome based on functional annotation of
60,770 full-length cDNAs.";
Nature 420:563-573(2002).
[4]
NUCLEOTIDE SEQUENCE.
STRAIN=C57BL/6J; TISSUE=Thymus;
MEDLINE=20493374; PubMed=11042159; DOI=10.1101/gr.145100;
Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
"Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes.";
Genome Res. 10:1617-1630(2000).
[5]
NUCLEOTIDE SEQUENCE.
STRAIN=C57BL/6J; TISSUE=Thymus;
MEDLINE=20530913; PubMed=11076661; DOI=10.1101/gr.152600;
Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
Fujiwara S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsushima S., Kawai J.,
Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
"RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer.";
Genome Res. 10:1757-1771(2000).
[6]
NUCLEOTIDE SEQUENCE.
STRAIN=C57BL/6J; TISSUE=Thymus;
Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
Nishi K., Nomura K., Numazaki R., Ono M., Ohsato N., Okazaki Y.,
Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
Sasaki D., Shibata K., Shingawa A., Shiraki T., Sogabe Y., Tagami M.,
Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
Tomaru A., Toyota T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
EMBL: AK080321; BAG37878.1; -, mRNA.
Ensembl: ENSMUSG00000030276; Mus musculus.
MGI: MGI:2141418; 4833441J24Rik.
GO: GO:0004835; F:ubulin-tyrosine ligase activity; IEA.
GO: GO:0006464; P:protein-tyrosination; IEA.

Qy 180 RSPLLPAHNNWTSRFOEYLQROGRGAVMGSVIYPSMKKAIAHAMKVAQDHVEPRKNSFE 239
Db 122 RHPFLPDNNWSSQKFOAHQLEGTGAPNWSVIVPGMKAIVIAHALQTSQDTVOCRASFE 181
Qy 240 LYGADFVGLGRDPRFWLIEINSSPTMHPSTPVTALQCAQVOEDTIKVAVDNRSCD 292
Db 182 LYGADFVFGEDFQFWLIEINASTPMTAPSTAVTARLCAGVQADTLRVVVIDWRD 234

RESULT 12
Q922T0 MOUSE
ID Q922T0 MOUSE PRELIMINARY; PRT; 266 AA.
AC Q922T0;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE 4833441J24Rik protein.
GN Names=4833441J24Rik;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=FVB/N; TISSUE=Mammary tumor. C3;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalek U., Smailus D.E.,
RA Schermer A., Schein J.E., Jones S.J.M., Marra M.A.; and
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=FVB/N; TISSUE=Mammary tumor. C3;
RG NIH MGC Project;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC006830; AA006830.1; -; mRNA.
DR Ensembl; ENSMUSG0000030276; Mus musculus.
DR MGI; MGI:2141418; 4833441J24Rik.
DR GO; GO:0004835; P: tubulin-cytosine ligase activity; IEA.
DR GO; GO:0006464; P: protein modification; IEA.
DR InterPro; IPR004344; Tub_tyr_ligase.
DR Pfam; PF03133; TTL; 1.
SQ SEQUENCE 266 AA; 30506 MW; F8E8FB52FA8B8E98 CRC64;

Query Match 51.6%; Score 805; DB 2; Length 266;
Best Local Similarity 61.0%; Pred. No. 2.1e-60;
Matches 144; Conservative 41; Mismatches 47; Indels 4; Gaps 1;

Qy 61 IDGLRNWITKPAKSRGRDIVCMRDVEILELAADHPILSRDNKVVQKVIETPLICD 120
Db 1 MEGQNIWIWAPGAASRGIVCMNRIDEMLKLVDCNPLMKQGVQKVIETPLIFG 60
Qy 121 TKFDIRQWFLVDNPLTIWFKESYLRFSTQRFSLDKLDSATHLCNNVQKYLKNDVGR 180
Db 61 TKFDLRQWFLVDNPLTIWFKESYLRFSTQRFSLDKLDSATHLCNNVQKYLKNDVGR 120

Qy 181 SPILLPAHNNWTSRFOEYLQROGRGAVMGSVIYPSMKKAIAHAMKVAQDHVEPRKNSFEL 240
Db 121 HPMFLPDNNWSSQKFOAHQLEVDAPKAWSSVIVPGMKAIVIAHALQTSQDNVQCRKASFEL 180
Qy 241 YGADFVGLGRDPRFWLIEINSSPTMHPSTPVTALQCAQVOEDTIKVAV----DRSCD 292
Db 181 YGADFVFGEDFQFWLIEINASTPMTAPSTAVTARLCAGVQADTLRVVVIDRDLRSCD 236

RESULT 13
Q7Q156 ANOGA
ID Q7Q156 ANOGA PRELIMINARY; PRT; 572 AA.
AC Q7Q156;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE ENSANGP0000022337 (Fragment).
GN ORFNames=ENSANG0000019848;
OS Anopheles gambiae str. PEST.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Culicidae;
OC Anophelinae; Anopheles.
OX NCBI_TaxID=180454;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=PEST;
RG The Anopheles gambiae Sequence Committee;
RT "Anopheles gambiae re-annotation."
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=PEST;
RG The Anopheles gambiae Sequence Committee;
RL Submitted (APR-2004) to the EMBL/GenBank/DBJ databases.
CC -1- CAUTION: The sequence shown here is derived from an
EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
preliminary data.
DR EMBL; ARA801008980; EAA13905.2; -; Genomic DNA.
DR GO; GO:0004835; P: tubulin-cytosine ligase activity; IEA.
DR GO; GO:0006464; P: protein modification; IEA.
DR InterPro; IPR004344; Tub_tyr_ligase.
DR Pfam; PF03133; TTL; 1.
FT NON_TER 1 1
FT NON_TER 572 572
SQ SEQUENCE 572 AA; 67109 MW; 8938D9DD5935071 CRC64;

Query Match 43.1%; Score 673; DB 2; Length 572;
Best Local Similarity 45.0%; Pred. No. 1.1e-48;
Matches 135; Conservative 54; Mismatches 97; Indels 14; Gaps 5;

Qy 2 DIDSADAVEDLTEAEWEDLTQYYSLVHGD---AFISNSR-----NYFSOCCALLNRI 52
Db 239 DIDIEEDT--KVDHDDWDFVLTTHYLLTHEDNRITQLKEEREADAIEHYLAESKSVLEQI 296
Qy 53 TSNVPOTDIDGLRNWITKPAKSRGRDIVCMRDVEILELAADHPILSRDNKVVQKVI 112
Db 297 KSHWPQYALDGLNINWIKVGNKCRGIIHLMNNIKQI--IAMVNPVIVSKTRVYQKI 354
Qy 113 ETPLLICDTKFDIRQWFLVDNPLTIWFKESYLRFSTQRFSLDKLDSATHLCNNVQK 172
Db 355 ERPLIHTKFDIRQWFLVDNPLTIWFKESYLRFSSQQYNLMNYHESVHLTHAIQK 414
Qy 173 YLKNVDCRSPLLPANNNMTSTRFOEYLQROGRGAVMGSVIYPSMKKAIAHAMKVAQDHVE 232
Db 415 KYHNAV-RDRLPHENNMDCHTFOAYLRQIDKVMWSEIRIYPMQKAIIGSLACQNMD 473
Qy 233 PRKNSFELYGADFVGLGRDPRFWLIEINSSPTMHPSTPVTALQCAQVOEDTIKVAVDNRSCD 292
Db 474 RRPNTFELYGADFVGLGRDPRFWLIEINSSPTMHPSTPVTALQCAQVOEDTIKVAVDNRSCD 533

RESULT 14
Q9VM91_DROME

ID Q9VM91_DROME PRELIMINARY; PRT; 992 AA.
 AC Q9VM91;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-OCT-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE CG11323-PA.
 GN Name=CG11323; ORFNames=CG11323;
 OS Drosophila melanogaster (fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Brill J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burris K.C., Busam D.A., Butler H., Cadiou E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dunham-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin C.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fostel C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lal Z.,
 RA Lasok P., Lei Y., Levitsky A.C., Li J.H., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195 (2000).
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RX MEDLINE=22426065; PubMed=12537568;
 RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
 RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
 RA George R.A., Hoskins R.A., Lavery T., Muzny D.M., Nelson C.R.,
 RA Pacle J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
 RA Svirskaas R., Taber P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
 RA Weinstein G., Scher S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
 RT "Finishing a whole-genome shotgun: release 3 of the *Drosophila*
 RT melanogaster euchromatic genome sequence.";
 RL Genome Biol. 3:RESEARCH0073-RESEARCH0079 (2002).
 RN [3]
 RP NUCLEOTIDE SEQUENCE.
 RX MEDLINE=22426070; PubMed=12537573;
 RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J.W., Svirskaas R.,
 RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
 RA Ashburner M., Celniker S.E.;
 RT "The transposable elements of the *Drosophila melanogaster* euchromatin:
 RT a genomics perspective."

RL Genome Biol. 3:RESEARCH0084.1-RESEARCH0084.20 (2002).
 RN [4]
 RP NUCLEOTIDE SEQUENCE.
 RX MEDLINE=22426069; PubMed=12537572;
 RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
 RA Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.O.,
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
 RA Lewis S.E.;
 RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a
 RT systematic review.";
 RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22 (2002).
 RN [5]
 RP NUCLEOTIDE SEQUENCE.
 RG Berkeley *Drosophila* Genome Project;
 RA Celniker S., Carlson J., Wan K., Pfeiffer B., Frise E., George R.,
 RA Hoskins R., Stapleton M., Pacle J., Park S., Svirskaas R., Smith E.,
 RA Yu C., Rubin G.;
 RT "Drosophila melanogaster release 4 sequence.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP NUCLEOTIDE SEQUENCE.
 RG FlyBase;
 RL Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AE003614; AAF52432.1; -; Genomic DNA.
 DR Ensembl; CG11323; *Drosophila melanogaster*.
 DR FlyBase; FBgn0031854; CG11323.
 DR GO; GO:0004835; F: tubulin-tyrosine ligase activity; IEA.
 DR GO; GO:0006464; P: protein modification; IEA.
 DR InterPro; IPR004344; Tub_tyr_lygase.
 DR Pfam; PF03133; TTL; 1.
 SQ SEQUENCE 992 AA; 112255 MW; 0FC34674C15BDAAD CRC64;
 Query Match 40.8%; Score 636; DB 2; Length 992;
 Best Local Similarity 42.8%; Pred. No. 3e-45;
 Matches 128; Conservative 59; Mismatches 96; Indels 16; Gaps 5;
 QY 2 DIDSADAVE--LITEAEWDLTQYYSVLVHGDAFISN-----RNYPSCQALLNRIT 53
 DB 349 DID-----FEDPPKIWEHDWDAFUFHQOOLVNEGDQIOHQDGGORLEPVMVSKSLVDKMK 403
 QY 54 SVNPQTDIDGLRNIIKPAKRSGRDVCMDRVEEILELAADHPLSRDNKWWVQKXIE 113
 DB 404 VHMQYSLDGYQNNWIKVANKRGRIILMDNLKLI--LGVNLSIASKRYVQKXIE 461
 QY 114 TPLICDTKFDIROWFLVTDNPLTIWPKSYLRFSQFSLDKLSDIAHLCNNVQKY 173
 DB 462 RPLILFQTKFDIROWFLITNTQPLVWFWFRESYLRFSQSYLSLNHSHVLTNYAIQKK 521
 QY 174 LKNDVGRSPLPAHNMWTSFTFOYLQRCRGVWGSVYFYSMKKAIHAHMKVAQDVPEP 233
 DB 522 YTNQ-KDKRLPSPNNMMDVCSFYQALYRQIGKYNMMLRIFPFGRKAIYVGCMLASQENMDR 580
 QY 234 RKNSELYGADFVLGRDPRMPLIENSPTVHPSTVTAQLCAQVOEDTIKVAVDKSD 292
 DB 581 RPNTFELFGDFMICEFYPLIENSFDLGATTSVTARMCPQCLDVKVVDIRRTD 639
 RESULT 15
 ID Q7PMD3 ANOGA PRELIMINARY; PRT; 501 AA.
 AC Q7PMD3;
 DT 01-MAR-2004 (TrEMBLrel. 26, Created)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
 DE ENSANGP0000022334 (Fragment).
 GN ORFNames=ENSANGG00000019845;
 OS Anopheles gambiae str. PEST.
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Culicidae;
 OC Anophelinae; Anopheles.

```
OX NCBI_TaxID=180454;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=PEST;
RG The Anopheles gambiae Sequence Committee;
RT "Anopheles gambiae re-annotation.";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=PEST;
RG The Anopheles gambiae Sequence Committee;
RL Submitted (APR-2004) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
preliminary data.
CC
DR EMBL; AAB01008980; EAA13971.3; -; Genomic DNA.
DR GO; GO:0004835; F:tubulin-tyrosine ligase activity; IEA.
DR GO; GO:0006464; P:protein modification; IEA.
DR InterPro; IPR011761; ATP_GRASP.
DR InterPro; IPR004344; Tub_tyr_lygase.
DR Pfam; PF03133; TTL; 1.
DR PROSITE; PS50975; ATP_GRASP; 1.
FT NON_TER 1
FT NON_TER 501
FT SEQUENCE 501 AA; 58862 MW; 806539A2FD8E76DF CRC64;

Query Match 34.6%; Score 539.5; DB 2; Length 501;
Best Local Similarity 41.1%; Pred. No. 2.3e-37;
Matches 123; Conservative 50; Mismatches 111; Indels 15; Gaps 5;

QY 1 EDIDTGADEDLTEAEWEDLTQOYSLVHGDAFINSRNY-----FSQOALLNRITSV 55
Db 201 EDIDHEEFCPLTPENRHEMEQ----LLAGEAQFGSIPGYEVLSLVAQYKLVHEAAKV 256

QY 56 NPQTDIDGLNMIKPAKSRGRDIVCMRDVREIELELAADHPLSRDNKMWVQKYETP 115
Db 257 FPMKIDGIRNMWILKPGNCRGLGIMLFNDDRKLLLEHVD-----NPDVKYVAQKYIERP 312

QY 116 LLICDTKFDIRQWFLVT-DWNPLTIWFKESYLRFSTQRFSLDKLPSAIHLCNNAVQK-Y 173
Db 313 LLIHCTKFDIRQYFLITYTNNVLKWMYRNCYLRFSSRQFNLDQFSEIHLTNYSIQKNY 372

QY 174 LKNDVGRSPLLPANNMTSTRFQBYLQROGRGAVGWSVIYPSMKKAIAHAMKVAQDHVEP 233
Db 373 AKEVREGADALPASNMWMSLKRQFQHLQSLDKGFYWERKIYPMKKNIILAIVCASLDGMKM 432

QY 234 RKNSEFELYGADFVLGRDFPWLIEINSSPTMHPSTPTVTAQLCAQVOEDTIKAVDRSCD 292
Db 433 ERNMFELYGADFVWTFNFRMLIEINISPDLSSTSDVTSTVTCPAVLEDLVKVVIDNTKD 491
```

Search completed: April 4, 2006, 12:45:57

Job time : 45.3857 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2006, 12:45:31 ; Search time 10.8922 Seconds
(without alignments)
2588.231 Million cell updates/sec

Title: US-10-635-977-2_COPY_73_365
Perfect score: 1560
Sequence: 1 EDIDTSADAVEDLTEAWED.....LCAQVEDTIKVAVDRCDI 293

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	908	58.2	352	2 T12515	hypothetical prote
2	255.5	16.4	379	2 A45443	tubulin-tyrosine 1
3	244	15.6	640	2 E89575	protein ZK1128.6 {
4	244	15.6	680	2 T27699	hypothetical prote
5	209.5	13.4	1198	2 T20262	hypothetical prote
6	209.5	13.4	1203	2 C93217	protein C55A6.2 {i
7	166	10.6	662	2 T20343	hypothetical prote
8	160.5	10.3	403	2 T37571	tubulin-tyrosine 1
9	155.5	10.0	883	2 A96805	hypothetical prote
10	144.5	9.3	753	2 S48261	hypothetical prote
11	95.5	6.1	2176	2 T39188	probable U5 snRNP-
12	95	6.1	1217	2 T21403	hypothetical prote
13	94.5	6.1	1528	2 S13743	DNA strand transfe
14	94	6.0	626	2 T49625	glucan 1,4-alpha-9
15	94	6.0	626	2 S36364	glucan 1,4-alpha-9
16	93	6.0	457	2 T29116	hypothetical prote
17	91.5	5.9	4660	2 T42737	gp330 protein prec
18	90.5	5.8	816	2 T19049	hypothetical prote
19	89	5.7	2870	2 H96974	cyclic beta 1-2 gl
20	88	5.6	282	2 E84317	hypothetical prote
21	87.5	5.6	1270	2 T22615	hypothetical prote
22	87.5	5.6	3971	2 T44806	myosubtilin synth
23	87.5	5.6	4688	2 F82885	hypothetical prote
24	87	5.6	733	2 E71937	flagellar biosynth
25	87	5.6	1044	2 S16516	integrin alpha-8 c
26	86.5	5.5	733	2 A64650	flagellar biosynth
27	86.5	5.5	796	2 T16367	hypothetical prote
28	86.5	5.5	1857	1 S01787	fatty-acid synthas
29	86	5.5	295	2 T30703	probable mRNA guan

30	85.5	5.5	1124	2 JH0588	calmodulin-binding
31	85	5.4	590	2 H70130	oligoendopeptidase
32	84.5	5.4	445	2 I46225	mucin A, tracheal
33	84.5	5.4	522	2 T45824	hypothetical prote
34	84	5.4	439	2 T01270	hypothetical prote
35	84	5.4	1872	2 T30888	vitellogenin - Ath
36	84	5.4	2182	2 T14320	calcineurin inhibi
37	83.5	5.4	930	2 A25923	progesterone recep
38	83	5.3	4128	2 JC6306	protein kinase (EC
39	82.5	5.3	275	2 G75130	translation initia
40	82.5	5.3	349	2 F91218	probable transport
41	82.5	5.3	349	2 H86064	probable transport
42	82.5	5.3	349	2 D65182	probable transport
43	82.5	5.3	396	2 T04561	hypothetical prote
44	82.5	5.3	609	2 B84783	probable poly(A) b
45	82.5	5.3	1297	2 T52065	probable myb-relat

ALIGNMENTS

RESULT 1

T12515
hypothetical protein DKFp434B103.1 - human
C:Species: Homo sapiens (man)
C>Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 09-Jul-2004
C:Accession: T12515
R:Wambutt, R.; Heubner, D.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, June 1999
A:Reference number: Z17524
A:Accession: T12515
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-352 <WAM>
A:Cross-references: UNIPROT:O9Y4R7; UNIPARC:UPI00001377A6; EMBL:AL096725
A:Experimental source: adult testis; clone DKFp434B103
C:Genetics:
A:Note: DKFp434B103.1

Query Match	58.2%	Score	908	DB	2	Length	352
Best Local Similarity	56.7%	Pred.	No. 7.2e-72				
Matches	165	Conservative	48	Mismatches	78	Indels	0
Gaps	0						
Qy	2	DIDTSADAVEDLTEAWEDLTQQYSLVHGDAFISNSRNFSCQALLNRITSVNPQTDI	61				
Db	2	DIDKDLAPLYLTPEGWSLFLQRYQVVEGAEHLDTQVRCEDILQQLQAVVPQIDM	61				
Qy	62	DGLRNIIKPAAKSRGRDIVCMORVEILEAAADHPLSRDNKRWVQKYIETPLLICDT	121				
Db	62	EGDRNIWIVKPGAKSRGRGIMCMHLEMLKLVNGNPVVMKDGKVVQKYIERPLLI	121				
Qy	122	KFDIRQWFLVTDWNPPLTTFVFKESYLRESTQRFSLDKLDSAIHLCCNNAVOKYLKNDVGRS	181				
Db	122	KFDLRQWFLVTDWNPPLTTFVYRDSYIFRSTQFPLSLKNDNSVHLCCNSIQHLENSCHRH	181				
Qy	182	PLLPANHWMTSTRFOYLQROGRGAVMGSVTPYSMKKAIAMKVAQDHVEPRKNSFELY	241				
Db	182	PLLPDPNNWSQRQAHLENGAPNANSTIIVPGMKDAVIALQTSQDTVQCRKASFELY	241				
Qy	242	GADFVLGRDFRPWLIEINSSPTMHPSTPTVTAQLCAQVEDTIKVAVDRCSD	292				
Db	242	GADFVFGDFQFWLIEINASTMAPSTAVTARLCAGVQADTLRVVIDRMLD	292				

RESULT 2

A45443
tubulin-tyrosine ligase (EC 6.3.2.25) [validated] - pig
C:Species: Sus scrofa domestica (domestic pig)
C>Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
C:Accession: A45443
R:Ersfeld, K.; Wehland, J.; Plessmann, U.; Dodemont, H.; Gerke, V.; Weber, K.
J. Cell Biol. 120, 725-732, 1993
A:Title: Characterization of the tubulin-tyrosine ligase.

A:Reference number: A45443; MUID:93147125; PMID:8093886
A:Accession: A45443
A>Status: preliminary
A:Molecule type: mRNA; protein
A:Residues: 1-379 <ERS>
A:Cross-references: UNIPROT:P38160; UNIPARC:UPI00001377A9; GB:X68453; GB:S54050; NID:921
A:Experimental source: brain
A>Note: sequence extracted from NCBI backbone (NCBIN:123858, NCBIF:123859)
C:Keywords: ligase

Query Match 16.4%; Score 255.5; DB 2; Length 379;
Best Local Similarity 28.4%; Pred. No. 1.3e-14;
Matches 74; Conservative 46; Mismatches 98; Indels 43; Gaps 11;

QY 36 SNSRNVFSQOALLNRTISVPTDIDGLRNIIWIKPAAKSRGRDIVCMRDVVEEILELAA 95
DB 125 TREFE-----LTSYNKKE-DGEGNWIAKSSAGAKGEGILISEATELDDFI- 173
QY 96 ADHPLSRDNK---WVVKYIETPLI--CDTKFDIQWFLVTDWNPITTFWFKESYLRS 150
DB 174 -----DNQGVHVIOKYLRLPDLLEPGHRKFDIRSVLVD--HOYNIYLYREGVLRTA 224
QY 151 TQFSLDKI-DSAIHLGNNAVKYKNDVGRSPLLPANNNWTSRFEYLQQRGAVWG 209
DB 225 SEPYHTDNFQDKTCHLTNHCIOKEYSKNYGK---YEGNEMFFEEFNQYLT-----SALN 276
QY 210 SVIYPSMKKAIAHAMVAQDHVPR-----KNSFELYGADPVLGRDFRPMLEINSSP 262
DB 277 ITUESILLOIKHIIKRSCLLSVPAISTRLHPYQSOLFQGFDFMVDLKWLLIEVNGAP 336
QY 263 TMHPSTPVTQAQCAQVQEDTI 283
DB 337 AC--AQKLYAELCOGIVDIAI 355

RESULT 3
E88575
protein ZK1128.6 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
C:Accession: E88575
R:Anonymous, The C. elegans Sequencing Consortium.
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
S:Science 282, 2012-2018, 1998
A:Reference number: A75000; MUID:99069613; PMID:9851916
A>Note: see websites genome.wuatl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C_ele
A>Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A:Accession: E88575
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-640 <STO>
A:Cross-references: UNIPROT:Q09647; UNIPARC:UPI000017A077; GB:chr_III; PIDN:CAA87425.1;
C:Genetics:
A:Gene: ZK1128.6
A:Map position: 3

Query Match 15.6%; Score 244; DB 2; Length 640;
Best Local Similarity 27.6%; Pred. No. 2.7e-13;
Matches 67; Conservative 48; Mismatches 90; Indels 38; Gaps 9;

QY 60 DIDGLRNIIWIKPAAKSRGRDIVCMRDVVEEILELAAADHPLSRDNKVVQKYIETPLLI 119
DB 283 ETDASRHV-IVKPPASARGTGISVTRKPKDPTTATL-----VAQHYIERPLTIN 331
QY 120 DTKFDIQWFLVTDWNPITTFWFKESYLRSYLRSTQRFSLDK---LDSAIHLGNNAVKYKLN 176
DB 332 RAKFDLRLAYVPTFEPLRVYIDQGLVRFASVPYSHSVSTISNKYMHLTNYSINKLAE 391
QY 177 D-VGRSPL--LPAHNNWTSRFE-----OEYLQQRGAVWGVSIVYPSMKKAIAHAMKVA 227
DB 392 DGVANKVPKWTLLHLW--EHFDEMVDREKIQRE-----IEEVIKAFISTEKPI 440
QY 228 QDH-----VEPRKNSFELYGADPVLGRDFRPMLEINSSPTMHPSTPVTQAQCAQVQEDTI 283

A:Reference number: A45443; MUID:93147125; PMID:8093886
A:Accession: A45443
A>Status: preliminary
A:Molecule type: mRNA; protein
A:Residues: 1-379 <ERS>
A:Cross-references: UNIPROT:P38160; UNIPARC:UPI00001377A9; GB:X68453; GB:S54050; NID:921
A:Experimental source: brain
A>Note: sequence extracted from NCBI backbone (NCBIN:123858, NCBIF:123859)
C:Keywords: ligase

Query Match 16.4%; Score 255.5; DB 2; Length 379;
Best Local Similarity 28.4%; Pred. No. 1.3e-14;
Matches 74; Conservative 46; Mismatches 98; Indels 43; Gaps 11;

QY 36 SNSRNVFSQOALLNRTISVPTDIDGLRNIIWIKPAAKSRGRDIVCMRDVVEEILELAA 95
DB 125 TREFE-----LTSYNKKE-DGEGNWIAKSSAGAKGEGILISEATELDDFI- 173
QY 96 ADHPLSRDNK---WVVKYIETPLI--CDTKFDIQWFLVTDWNPITTFWFKESYLRS 150
DB 174 -----DNQGVHVIOKYLRLPDLLEPGHRKFDIRSVLVD--HOYNIYLYREGVLRTA 224
QY 151 TQFSLDKI-DSAIHLGNNAVKYKNDVGRSPLLPANNNWTSRFEYLQQRGAVWG 209
DB 225 SEPYHTDNFQDKTCHLTNHCIOKEYSKNYGK---YEGNEMFFEEFNQYLT-----SALN 276
QY 210 SVIYPSMKKAIAHAMVAQDHVPR-----KNSFELYGADPVLGRDFRPMLEINSSP 262
DB 277 ITUESILLOIKHIIKRSCLLSVPAISTRLHPYQSOLFQGFDFMVDLKWLLIEVNGAP 336
QY 263 TMHPSTPVTQAQCAQVQEDTI 283
DB 337 AC--AQKLYAELCOGIVDIAI 355

RESULT 3
E88575
protein ZK1128.6 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
C:Accession: E88575
R:Anonymous, The C. elegans Sequencing Consortium.
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
S:Science 282, 2012-2018, 1998
A:Reference number: A75000; MUID:99069613; PMID:9851916
A>Note: see websites genome.wuatl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C_ele
A>Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A:Accession: E88575
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-640 <STO>
A:Cross-references: UNIPROT:Q09647; UNIPARC:UPI000017A077; GB:chr_III; PIDN:CAA87425.1;
C:Genetics:
A:Gene: ZK1128.6
A:Map position: 3

Query Match 15.6%; Score 244; DB 2; Length 640;
Best Local Similarity 27.6%; Pred. No. 2.7e-13;
Matches 67; Conservative 48; Mismatches 90; Indels 38; Gaps 9;

QY 60 DIDGLRNIIWIKPAAKSRGRDIVCMRDVVEEILELAAADHPLSRDNKVVQKYIETPLLI 119
DB 283 ETDASRHV-IVKPPASARGTGISVTRKPKDPTTATL-----VAQHYIERPLTIN 331
QY 120 DTKFDIQWFLVTDWNPITTFWFKESYLRSYLRSTQRFSLDK---LDSAIHLGNNAVKYKLN 176
DB 332 RAKFDLRLAYVPTFEPLRVYIDQGLVRFASVPYSHSVSTISNKYMHLTNYSINKLAE 391
QY 177 D-VGRSPL--LPAHNNWTSRFE-----OEYLQQRGAVWGVSIVYPSMKKAIAHAMKVA 227
DB 392 DGVANKVPKWTLLHLW--EHFDEMVDREKIQRE-----IEEVIKAFISTEKPI 440
QY 228 QDH-----VEPRKNSFELYGADPVLGRDFRPMLEINSSPTMHPSTPVTQAQCAQVQEDTI 283

Db 441 REHMSRFLQEFCYELFGIDILDEDYKFWLLEVNISPSLSHSGTPLDVSVKAPLAKDVL 500
QY 284 KVA 286
DB 501 NLA 503
RESULT 4
T20262
hypothetical protein ZK1128.6 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T20262
R:Berks, M.
submitted to the EMBL Data Library, January 1995
A:Reference number: Z20407
A:Accession: T20262
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-680 <WIL>
A:Cross-references: UNIPROT:Q09647; UNIPARC:UPI000017A076; EMBL:Z47357; PIDN:CAA87425.2
A:Experimental source: clone ZK1128
C:Genetics:
A:Gene: CESP:ZK1128.6
A:Map position: 3
A:Introns: 114/1; 208/3; 242/3; 334/3; 362/2; 405/2; 424/2; 550/2; 633/2

Query Match 15.6%; Score 244; DB 2; Length 680;
Best Local Similarity 27.6%; Pred. No. 2.9e-13;
Matches 67; Conservative 48; Mismatches 90; Indels 38; Gaps 9;

QY 60 DIDGLRNIIWIKPAAKSRGRDIVCMRDVVEEILELAAADHPLSRDNKVVQKYIETPLLI 119
DB 323 ETDASRHV-IVKPPASARGTGISVTRKPKDPTTATL-----VAQHYIERPLTIN 371
QY 120 DTKFDIQWFLVTDWNPITTFWFKESYLRSYLRSTQRFSLDK---LDSAIHLGNNAVKYKLN 176
DB 372 RAKFDLRLAYVPTFEPLRVYIDQGLVRFASVPYSHSVSTISNKYMHLTNYSINKLAE 431
QY 177 D-VGRSPL--LPAHNNWTSRFE-----OEYLQQRGAVWGVSIVYPSMKKAIAHAMKVA 227
DB 432 DGVANKVPKWTLLHLW--EHFDEMVDREKIQRE-----IEEVIKAFISTEKPI 480
QY 228 QDH-----VEPRKNSFELYGADPVLGRDFRPMLEINSSPTMHPSTPVTQAQCAQVQEDTI 283
DB 481 REHMSRFLQEFCYELFGIDILDEDYKFWLLEVNISPSLSHSGTPLDVSVKAPLAKDVL 540
QY 284 KVA 286
DB 541 NLA 543
RESULT 5
T20262
hypothetical protein C55A6.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T20262
R:Kershaw, J.
submitted to the EMBL Data Library, October 1996
A:Reference number: Z19243
A:Accession: T20262
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1198 <WIL>
A:Cross-references: UNIPROT:O17720; UNIPARC:UPI000017A066; EMBL:Z81051; PIDN:CAB02862.2
A:Experimental source: clone C55A6
C:Genetics:
A:Gene: CESP:C55A6.2
A:Map position: 5
A:Introns: 32/3; 68/1; 107/3; 201/3; 273/1; 364/3; 407/3; 460/3; 476/2; 534/2; 677/3; 7

Query Match 13.4%; Score 209.5; DB 2; Length 1198;
Best Local Similarity 25.5%; Pred. No. 6.7e-10;
Matches 60; Conservative 57; Mismatches 87; Indels 31; Gaps 9;
QY 68 WIIKPAKSRGRDIVCMRVEILEAAADHPLSRDNKMWVQKYIETPLLCIDTKFDIRQ 127
DB 715 FIVKPTNSRQKGIFFANSMADI-----PAGGL-----LVSRYLKDPYLVNNHKFDLRI 764
QY 128 WFLVTDNPLTIWFKYKSYLRFSTQRF--SLDKLDS--AIHLCNNAVOK---YLNKNDVG 179
DB 765 YVAVTSFYPLAVYVSEGLARLASRPYDTSASSADSNVHLTNYSINKNSTSFVRNESH 824
QY 180 RSPLLPAHNMWTSRFOYLQGRGAVGWSVIYPSMKKAIHAHMKVAQDHV--EPRKQ- 236
DB 825 SSEDLL--GHKWTGLALLRYVNEGKDA---KILMLRIEDLVKSLLSIQNSVATASRTNL 879
QY 237 -----SFELYGADFVLGRDPRFWLIEINSPMTWHPSTPTVTAQLCAQVQEDTIKVA 286
DB 880 RFACTNFELFGFDVLVDQALKPWLLEVNLSPLACDAPLDSLLKTRLIADLLNLA 934
RESULT 6
C89217
protein C55A6.2 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
C:Accession: C89217
R:Anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
A:Reference number: AY5000; PMID:9906913; PMID:9851916
A:Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C.ele
A:Accession: C89217
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1203 <STO>
A:Cross-references: UNIPROT:O17720; UNIPARC:UPI000017A067; GB:chr_V; PIDN:CAB02862.1; PI
C:Genetics:
A:Map position: 5
Query Match 13.4%; Score 209.5; DB 2; Length 1203;
Best Local Similarity 25.5%; Pred. No. 6.7e-10;
Matches 60; Conservative 57; Mismatches 87; Indels 31; Gaps 9;
QY 68 WIIKPAKSRGRDIVCMRVEILEAAADHPLSRDNKMWVQKYIETPLLCIDTKFDIRQ 127
DB 720 FIVKPTNSRQKGIFFANSMADI-----PAGGL-----LVSRYLKDPYLVNNHKFDLRI 769
QY 128 WFLVTDNPLTIWFKYKSYLRFSTQRF--SLDKLDS--AIHLCNNAVOK---YLNKNDVG 179
DB 770 YVAVTSFYPLAVYVSEGLARLASRPYDTSASSADSNVHLTNYSINKNSTSFVRNESH 829
QY 180 RSPLLPAHNMWTSRFOYLQGRGAVGWSVIYPSMKKAIHAHMKVAQDHV--EPRKQ- 236
DB 830 SSEDLL--GHKWTGLALLRYVNEGKDA---KILMLRIEDLVKSLLSIQNSVATASRTNL 884
QY 237 -----SFELYGADFVLGRDPRFWLIEINSPMTWHPSTPTVTAQLCAQVQEDTIKVA 286
DB 885 RFACTNFELFGFDVLVDQALKPWLLEVNLSPLACDAPLDSLLKTRLIADLLNLA 939
RESULT 7
T20343
hypothetical protein D2013.9 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T20343; T22085
R:Mortimore, B.
submitted to the EMBL Data Library, January 1995
A:Reference number: Z19259
A:Accession: T20343

A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-662 <WIL>
A:Cross-references: UNIPROT:O09512; UNIPARC:UPI000013B847; EMBL:Z47808; PIDN:CAA87778.
A:Experimental source: clone D2013
R:Matthews, P.
submitted to the EMBL Data Library, January 1995
A:Reference number: Z19510
A:Accession: T22085
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-662 <WIL>
A:Cross-references: UNIPARC:UPI000013B847; EMBL:Z47809; PIDN:CAA87783.1; GSPDB:GN00020
A:Experimental source: clone F42A8
C:Genetics:
A:Gene: CESP:D2013.9
A:Map position: 2
A:Introns: 78/2; 114/3; 330/3; 471/1; 521/2; 587/2
Query Match 10.6%; Score 166; DB 2; Length 662;
Best Local Similarity 24.3%; Pred. No. 2e-06;
Matches 72; Conservative 45; Mismatches 115; Indels 64; Gaps 14;
QY 6 SADAVEDLTEAEWDLTQOYVSL-VHGDAFISNSRNYFSQCALLNRITSVNPQTDIDGL 64
DB 396 AACAMRDPKNDWYQLT---YLNATQLPEFVARFQN-----RELNGQ 434
QY 65 RNIIWIKPAKSRGRDIVCMRVEILEAAADHPLSRDNKMWVQKYIETPLLI----- 118
DB 435 HNVILVKPNWNLARGMDMTVTDNLNLIIRI-----ETGPKIVICEYIPRLLFRPPDNG 487
QY 119 CDTKFDIRQWFLVTDNPLTIWFKYKSYLRFSTQRFSLDKL-DSAIHLCNNAVQYKLNKD 177
DB 488 NKVFDLRYIVFLNGIAPVTAVVYRWFIRAINFEFSLNPFEDVETHF---TVFNYL--D 542
QY 178 VGRSPLPAHNMWTSRFOYLQGRGAVGWSV---IYPSMKKAI-AHAMKVAQDHVEP 233
DB 543 KEKILQMKCN-----FIETIKAYPRIQWSEVQKIDNLTIRKIAEAAAKEAPRGVAP 596
QY 234 RKNSELYGADFVLGRD---FRPWLIEINSPMTWHPSTPTVTAQLCAQVQ--DTI 283
DB 597 NVQGRAMYGVDMILQHGNDVIKSTLLEINFPD-----TTRACQYYPDPADTV 645
RESULT 8
T37571
tubulin-tyrosine ligase related protein - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C:Accession: T37571
R:Badcock, K.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
submitted to the EMBL Data Library, April 1996
A:Reference number: Z21726
A:Accession: T37571
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-403 <BAD>
A:Cross-references: UNIPROT:Q10438; UNIPARC:UPI000013A94B; EMBL:Z70721; PIDN:CAA94694.
A:Experimental source: strain 972h-; cosmid c12B10
C:Genetics:
A:Gene: SPDB:SPAC12B10.04
A:Map position: 1
A:Introns: 320/3; 348/3
Query Match 10.3%; Score 160.5; DB 2; Length 403;
Best Local Similarity 20.5%; Pred. No. 3e-06;
Matches 63; Conservative 61; Mismatches 115; Indels 69; Gaps 12;
QY 26 YSLVHGDAFISNS-----RNYFSQCALLNRITSVNPQTDIDGLRNI-----WIILKP 72
DB 89 YLAKHPDSILKSVPEAYSLDELYAEFLDSDLSMEAYLRQLEENATKNISEKOWIILKP 148
QY 73 AAKSRGRDIVCMRVEILEL-----AAADHPLSRDNK-----WV 107


```

Best Local Similarity 19.1%; Pred.No.15;
Matches 65; Conservative 55; Mismatches 120; Indels 101; Gaps 15;

Qy 2 DIDSADAVEDLITE---AEWEDLT-----QQYSLVHGDAFISGRNYPSCQALLN 50
   :|::||::||::||::||::||::||::||::||::||::||::||::||::||:
Db 236 EADBEAAVEAMEEDEVLEDVLETSIQBEEKNIENPTEVT----FIADTKKVT 291
   :|::||::||::||::||::||::||::||::||::||::||::||::||:

Qy 51 RITSVNPTQDIDGLRNIIWIIPKAASRGSDIVCMRDVRVEILEAAADHPLSRDNKWVQK 110
   :|::||::||::||::||::||::||::||::||::||::||::||::||:
Db 292 EIPTHPR-EIDA---FWLQREIAIFYADVAVCOEKTNQAFAESALSDYDLGE----- 339
   :|::||::||::||::||::||::||::||::||::||::||::||::||:

Qy 111 YIETPLLICDTKFDIRQWFLV---TDWNPLTIWFYK-----ESYLRFSTQRF 154
   :|::||::||::||::||::||::||::||::||::||::||::||::||:
Db 340 -LENELM---SIFDYEHFLVLQLLTKNRNTIIVSCWTMLKRAATDEERGLGVEEQIRAGRSW 395
   :|::||::||::||::||::||::||::||::||::||::||::||::||:

Qy 155 SLDKL-----DSAIHLCHNNVQYKLYKNVDGRSPL-----LPAHNNMTSTRFQ 196
   :|::||::||::||::||::||::||::||::||::||::||::||::||:
Db 396 ILEALRPGAITIPDGLNELNANNVVEKEAPVPSEIPLSKTLTSHKIIVKHQV---DLE 451
   :|::||::||::||::||::||::||::||::||::||::||::||::||:

Qy 197 BYLQRQRGAVNGSVIYPSMKKAIAMKVAAQDHVEPRKNSFELYGADF----- 245
   :|::||::||::||::||::||::||::||::||::||::||::||::||:
Db 452 NYVFTEG-----SRLMSNKAVKLPESFRRTKGVEEIHVPAPNKA 492
   :|::||::||::||::||::||::||::||::||::||::||::||::||:

Qy 246 VLGRDFRPWLJEINSSPTMPHPSTPVTAQLCAOVQEDTIKVA 286
   ||||::||::||::||::||::||::||::||::||::||::||::||::||:
Db 493 VLGADER--LVKIKELPEWSHOAFLTQSLNRIQSHLYPIA 531
   ||||::||::||::||::||::||::||::||::||::||::||::||::||:

RESULT 12
T21403
hypothetical protein F26D2.10 - Caenorhabditis elegans
C;Species:Caenorhabditis elegans
C;Date:15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T21403
R;McMurray, A.
submitted to the EMBL Data Library, November 1996
A;Reference number: Z19418
A;Accession: T21403
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1217 <MWL>
A;Cross-references: UNIPROT:Q9XV62; UNIPARC:UPI0000060FB4; EMBL:Z81513; PIDN:CABO
A;Experimental source: clone F26D2
C;Genetics:
A;Gene: CESP:F26D2.10
A;Map position: 5
A;Introns: 29/1; 568/2; 608/3; 746/1; 871/3; 910/1; 985/1; 1018/3; 1124/1; 1170/3
C;Superfamily: Caenorhabditis elegans hypothetical protein F36D3.5

Query Match 6.1%; Score 95; DB 2; Length 1217;
Best Local Similarity 20.6%; Pred.No. 7.5;
Matches 58; Conservative 49; Mismatches 106; Indels 68; Gaps 12;

Qy 3 IDTSADAVEDLTEAWEDLTQQYSLVHGDAFISGRN-----YFSQ 44
   :|::||::||::||::||::||::||::||::||::||::||::||::||:
Db 78 LNTVVNVNVEVASSELRRMHQEYTKLIH---FFENRTOSSFNDDVNAIFNNIIFILYSR 134
   :|::||::||::||::||::||::||::||::||::||::||::||::||:

Qy 45 QQALLNRITSVNPTQDIDGLRNIIWIIPKAASRGSDIVC--MDRVEEILEAAADH--PL 100
   :|::||::||::||::||::||::||::||::||::||::||::||::||:
Db 135 KSSLISEVSSISTKNLANNETKS1-----AKDRLFNSCSLLTGLDEFVALFDQDLPN 188
   :|::||::||::||::||::||::||::||::||::||::||::||::||:

Qy 101 SRDNKKVVQXYIETPLICD-----TKED-----IROWFLYTDMNPITWFKESYLRF 150
   :|::||::||::||::||::||::||::||::||::||::||::||::||:
Db 189 SERKIAIINFHRQASLMTECISGLUKLDEIDALSNWKNVAYNKL--WYVKDSIILFE 246
   :|::||::||::||::||::||::||::||::||::||::||::||::||:

Qy 151 TORFSLDKLASIAHLCHNNVQYKLYKNVDGRSPLLPAHNNMTSTRFOEYLRQGRGAVNGS 210
   :|::||::||::||::||::||::||::||::||::||::||::||::||:
Db 247 TFREPLSS-----WNESLLKIL-----SAFAQTKPLWNS-----RSSIGQS 282
   :|::||::||::||::||::||::||::||::||::||::||::||::||:

Qy 211 VIYPSMKKAIAMKVAAQDHVEPRKNSFELYGADFVLGRDF 251
   :|::||::||::||::||::||::||::||::||::||::||::||::||:
Db 283 KIFEAMAQ-LNKSIVEADEHNMISSINDOLKYTAGFKNFCD 322
   :|::||::||::||::||::||::||::||::||::||::||::||::||:

```

RESULT 13

DNA strand transfrase 2 - yeast (Saccharomyces cerevisiae)
S13743
N:Alternate names: DST2 protein; KEM1 protein; protein YGL173c; RARS5 PR
C:Species: Saccharomyces cerevisiae
C>Date: 16-Sep-1992 #sequence revision 16-Sep-1992 #text change 09-Jul-2004
C:Accession: S13743; S16701; S16885; S42156; S52457; A39790; S59238; S64190
R:Kim, J.; Ljungdahl, P.O.; Fink, G.R.
Genetics 126, 799-812, 1990
A:Title: kem mutations affect nuclear fusion in Saccharomyces cerevisiae.
A:Reference number: S13743; MUID:91169260; PMID:2076815
A:Accession: S13743
A:Molecule type: DNA
A:Residues: 1-1528 <IM>
A:Cross-references: UNIPROT:P22147; UNIPARC:UPI000012DD97; EMBL:X54717; NID:g3846; PID
R:Tishkoff, D.X.; Johnson, A.W.; Kolodner, R.D.
Mol. Cell. Biol. 11, 2593-2608, 1991
A:Title: Molecular and genetic analysis of the gene encoding the Saccharomyces cerevis
A:Reference number: S16701; MUID:91203880; PMID:1840632
A:Accession: S16701
A:Molecule type: DNA
A:Residues: 1-1528 <IM>
A:Cross-references: UNIPARC:UPI000012DD97; GB:M58367; NID:g172579; PIDN:AAA35036.1; PI
R:Kipling, D.; Tambini, C.; Kearsey, S.E.
Nucleic Acids Res. 19, 1385-1391, 1991
A:Title: Rar mutations which increase artificial chromosome stability in Saccharomyces
A:Reference number: S16885; MUID:91227124; PMID:2027746
A:Accession: S16885
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1528 <IP>
A:Cross-references: UNIPARC:UPI000012DD97; EMBL:X61181; NID:g4412; PIDN:CAA34387.1; PI
R:Larimer, F.W.; Stevens, A.
submitted to the EMBL Data Library, March 1992
A:Description: Structure of the XRM1 exoribonuclease gene of Saccharomyces cerevisiae.
A:Reference number: S42156
A:Accession: S42156
A:Molecule type: DNA
A:Residues: 1-1528 <AR>
A:Cross-references: UNIPARC:UPI000012DD97; EMBL:M90097; NID:g173191; PIDN:AAA35219.1;
R:Bertani, I.; Coglievina, M.; Zaccaria, P.; Klima, R.; Bruschi, C.V.
submitted to the EMBL Data Library, February 1995
A:Description: The sequence of 11.1kb fragment on the left arm to Saccharomyces cerevi
A:Reference number: S52454
A:Accession: S52454
A:Molecule type: DNA
A:Residues: 1-1528 <BER>
A:Cross-references: UNIPARC:UPI000012DD97; EMBL:X84705; NID:g677853; PIDN:CAA59180.1; I
R:Dykstra, C.O.; Kitada, K.; Clark, A.B.; Hamatake, R.K.; Sugino, A.
Mol. Cell. Biol. 11, 2583-2592, 1991
A:Title: Cloning and characterization of DST2, the gene for DNA strand transfer protei
A:Reference number: A39790; MUID:91203879; PMID:1850100
A:Accession: A39790
A:Molecule type: DNA
A:Residues: 1-1260, 1', 1262-1528 <DYK>
A:Cross-references: UNIPARC:UPI0000178F97; GB:M36725
R:Bertani, I.; Coglievina, M.; Zaccaria, P.; Klima, R.; Bruschi, C.V.
Yeast 11, 1187-1194, 1995
A:Title: The sequence of an 11.1 kb fragment on the left arm of Saccharomyces cerevisi
A:Reference number: S59235; MUID:96109931; PMID:8619317
A:Accession: S59238
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1528 <BEW>
A:Cross-references: UNIPARC:UPI000012DD97; EMBL:X84705; NID:g677853; PIDN:CAA59180.1; I
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1995
R:Bruschi, C.V.; Coglievina, M.; Bertani, I.; Klima, R.; Zaccaria, P.; Delneri, D.
submitted to the Protein Sequence Database, May 1996
A:Reference number: S64183
A:Accession: S64190
A:Molecule type: DNA

A;Residues: 1-1528 <BRU>
A;Cross-references: UNIPARC:UPI000012D097; EMBL:Z72695; NID:g1322777; PIDN:CAA96885.1; F
A;Experimental source: strain S288C

C;Genetics:
A;Gene: SGD:KEM1; DST2; SEP1; RAR5; XRN1
A;Cross-references: SGD:S0003141; MIPS:YGL173C
A;Map position: 7L
C;Superfamily: Schizosaccharomyces pombe exonuclease II
C;Keywords: nucleus

Query Match	6.1%	Score 94.5;	DB 2;	Length 1528;
Best Local Similarity	20.5%;	Pred. No. 11;		
Matches 59:	Conservative	54;	Mismatches 132;	Indels 43;
				Gaps 13;

Qy	34	FINSRNYFQCQALLNRITSNP-QTDIGLKNIIWIKPAKSGRDIVCMORVEILLE	92
Dd	443	FITHSKSGS--YSLKMDDLSPDETETEEFQRVNSIRTKIKYQNATIVEDKEELETE	500
Qy	93	LAAADHPLSRDNKWVKQYIETPLLICDITKFDUR-----QMFLVTDMN--PLTIW	140
Dd	501	KTIYN---ERPERWKHEYHDKLKFTTDSSEKVRDLAKDYVEGLQWLVIYYVCPSWSW	557
Qy	141	FYKESYL-RSTQ-----RFSLDKLDGSAIHLCNNAVOKYILKNVDGRS--PLL----	184
Dd	558	YYPHHYAPRIADLAAGLDQDIEDLSKPFTFPQOLMAVLPERSKNLIPPAFRPLMTDEQS	617
Qy	185	PAHNMMWTSTRFOEYLORGGRGAWGVGSVIYISM--KKAIAHAMKVAQDHVPKRKNSPELYJ	242
Dd	618	PIHDFYFA---EVOLDKNGKTADWEAVLVISFVDEKELIEAMQPYLKRLSPEEKTRNQFG	674
Qy	243	ADFVLGRDFRPWLFIENSSPMTHPTPVTAQLCAQVOEDTIKVAVDRS	290
Dd	675	KDLIY--SFNPQVDNLYKSPLGGIFSIEDIENHC--VEKEYITTPLDSS	718

RESULT 14 -
T49625
glucan 1,4-alpha-glucosidase [imported] - Neurospora crassa
N:Alternate names: protein B5022.70
C:Species: Neurospora crassa
C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2003
C:Accession: T49625
R:Schulte, U.; Algin, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,
submitted to the Protein Sequence Database, May 2000
a:Reference number: 225022

A:Accession: J02522
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-626 <SCH>
A:Cross-references: UNIPARC:UPI000006A8CB; EMBL:ALJ55932; GSPDB:GN00116; NCSP:BSO22.70
A:Experimental source: BAC clone BSO22; strain OR74A
C:Genetics:
A:Gene: NCSP:BSO22.70

A-Introns: 82/1
A-map position: 6
C-Superfamily: glucan 1,4-alpha-glucosidase with starch-binding domain: glucan 1,4-alpha-

Query Match	6.0%;	Score 94;	DB 2;	Length 626;
Best Local Similarity	20.1%;	Pred. No. 3.7;		
Matches	72;	Conservative 40;	Mismatches 126;	Indels 120;
				Gaps 17;

QY 4 DTSADAVEDL TEAWEDLTQQYYSLVHGDAFISNRNYSQCQA----- 47

Db 207 NTGFDLWEYNSSSFFTTAASHRALVGSFAKSVSSSSCSACDIAIAPILCFQOQSWNS 266

Qy	48	--LLNRITSVNPQTIDG--LRNTWIIKPAAKSRGRDI-----VCMDRVEEILELAADHPL 100
	:	:
	:	:
Db	267	GVIISFNVYSRGKDINSVLTSIHNFOPAA---GCDWTEPQCSDR-----ALANKKV 316

Qy	101	SRDNK--W-----	VVOKYIETPLLCDTKFDQWFLVT-----	DWNPLTIW	140
			:	:	:
			:	:	:
Db	317	VVDSMRFGVNSGRTAGKAAAVGRYAE-----	DVYINGNPWYLATLAAAEQLYDAYVW		370

QY	141	-----FYKESYLRFSTQFSLDKLSAHLCCNNAVQKY	-----	171
DB	371	KKGQSIIVTSTSLAFFKDLVPSVSTGIYS--SSSSTYTAIIINAVTYYADGFDVIVAQYTP	428	
QY	174	-----LKNVGRSPLLPAHNWNTSTRFOEYLQROGRGAV---WGSVIYPSMKKATAH	222	
DB	429	SDGSLAEQFDRDSG-AFLSATHLTWYSASFLSAAARRA-GIVPPSWGSAANSILPGSCS-	485	
QY	223	AMKVAQDHVEPRKNSFELYGADFVLGRDFRPLWIEINSSPTMHPSPTPTAQLCAQVQE	280	
DB	486	ASTVAGSVATATATSF-----PANTLPASTTIVTPTQTGCAADHE	525	

RESULT 15

S36364
glucan 1,4-alpha-glucosidase (EC 3.2.1.3) precursor - *Neurospora crassa*
C:Alternate names: Glucoamylase; glycoamylase
C:Species: *Neurospora crassa*
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 02-Jun-2003
C:Accession: S36364; S13710; S13711; S25539
R:Stone, P.J.; Makoff, A.J.; Parish, J.H.; Radford, A.
Curr. Genet. 24, 205-211, 1993
A:Title: Cloning and sequence analysis of the glucoamylase gene of *Neurospora crassa*
A:Reference number: S36364; M1UD:94037144; PMID:8221928

A:Accession: S36384
A:Molecule type: DNA
A:Residues: 1-626 <STO>
A:Cross-references: UNIPARC:UPI000014A557; EMBL:X67291
R:Koh-Laur, S.I.; Parish, J.H.; Bleasby, A.J.; Pappin, D.J.C.; Ainley, K.; John-
Enzyme Microb. Technol. 11, 692-695, 1989
A:Title: Exported proteins of Neurospora crassa: 1-glucosylase.
A:Reference number: S13710
A:Accession: S13710
A:Molecule type: protein
A:Residues: 36-60, 'X', 62, 'X', 64-65 <KOH>
A:Cross-references: UNIPARC:UPI0000175A5C
R:Koh-Laur, S.I.; Parish, J.H.; Bleasby, A.J.; Pappin, D.J.C.; Ainley, K.; John-
submitted to the Protein Sequence Database, January 1990
A:Description: Exported proteins of Neurospora crassa 1: - glucosylase.

A:Reference number: 013711
A:Accession: S13711
A:Molecule type: protein
A:Residues: 36-60, 'X', 62, 'X', 64-65 <KOW>
A:Cross-references: UNIPARC:UPI0000175A5C
C:Genetics:
A:Gene: Gla-1
A:Introns: 82/2

C:Keywords: glycosidase; hydrolase; polysaccharide degradation
F1-19/Domain: signal sequence #status predicted <SIG>
F120-35/Domain: propeptide #status predicted <PRO>
F33-458/Domain: glucan 1,4-alpha-glucosidase homology <GAG>
F36-626/Product: glucan 1,4-alpha-glucosidase #status experimental <MAT>

Query Match	6.0%;	Score 94;	DB 2;	Length 626;
Best Local Similarity	20.1%;	Pred. No. 3.7;		
Matches	72;	Conservative 40;	Mismatches 126;	Indels 120;
Gaps	17			

Qy	4	D	T	S	A	D	A	V	E	L	T	E	A	E	W	E	L	T	Q	Y	S	L	V	H	G	D	A	F	I	S	N	S	N	Y	F	S	O	C	A	-----	47																		
Db	207	N	T	G	F	L	D	L	E	E	E	N	S	S	F	F	T	I	A	S	H	R	A	L	V	E	G	S	A	F	A	K	S	V	G	S	C	S	C	A	D	A	I	A	P	O	I	L	C	F	Q	Q	S	F	W	S	N	S	266

Qy	48	--LLNRITSVNPQTIDG--LNIWIWIKPAKSRGDI---VCMDRVEEILELAADHPL 100
Db	267	GYIIINFVYRSKGINSVLTSHNEDPAA---GCDVNTFQPCSDR-----ALANHKV 316

Qy	101	SRDNK-W-----VVQKYIETPLLCITKTDIRQWFLVT-----DWNPLTIW	140
Db	317	VVDSMRFWGVNSGRTAGKAAAVGRYAE-----DVTYGNPMYLATLAAAEQLYDAVVVV	370

141	Qy	-----FKYESYLRFSTQFSLDKLSAHLCNNAVQK-----	173
141	Qy	-----FKYESYLRFSTQFSLDKLSAHLCNNAVQK-----	173
371	Db	KKQGSITVSTSLAFFKDLVPSVSTGIS---SSSYHTAIINAVTTADGCFVDIVAQYTP	428

THIS PAGE BLANK (uspto)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2006, 12:42:35 ; Search time 93.4006 Seconds
(without alignments)
1378.343 Million cell updates/sec

Title: US-10-635-977-2_COPY_73_365

Perfect score: 1560

Sequence: 1 EIDITSADAVDLTAEWED.....LCAQVQEDTIKAVDRSDCI 293

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_21.*

- 1: Geneseqp1980s.*
- 2: Geneseqp1990s.*
- 3: Geneseqp2000s.*
- 4: Geneseqp2001s.*
- 5: Geneseqp2002s.*
- 6: Geneseqp2003as.*
- 7: Geneseqp2003bs.*
- 8: Geneseqp2004s.*
- 9: Geneseqp2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1560	100.0	293	8	ADJ93366 Human BGS
2	1560	100.0	541	8	ADJ93365 Human tub
3	1560	100.0	541	8	ADJ93358 Human BGS
4	1247	79.9	242	8	ADJ93360 Human BGS
5	908	58.2	352	3	AB43005 Human ORF
6	908	58.2	352	8	AAM39450 Human pol
7	908	58.2	352	8	ADJ93457 Human HOT
8	908	58.2	352	8	ABM80420 Tumour-as
9	908	58.2	399	6	ABU11512 Human MDD
10	908	58.2	488	5	AAU74334 Human cyt
11	908	58.2	744	4	AB94796 Human pro
12	905	58.0	326	7	ADM05524 Human pro
13	897	57.5	362	3	AB58909 Breast an
14	817	52.4	292	8	ADJ93455 Human HOT
15	636	40.8	992	4	AB65645 Drosophil
16	613	39.3	432	6	ABU92048 Human pro
17	515.5	33.0	719	4	AB65541 Drosophil
18	480	30.8	362	4	AAW41236 Human pol
19	439	28.1	281	8	ADG6614 Novel hum
20	374	24.0	566	4	ABG05971 Novel hum
21	336	21.5	160	8	ADH45424 Human mol
22	330	21.2	347	6	ABU00150 Human nov
23	324	20.8	496	4	AB64074 Drosophil
24	297.5	19.1	827	4	AB60840 Drosophil

25	296	19.0	1281	4	AAM39105 Human pol
26	296	19.0	1304	4	AAM40891 Human pol
27	290.5	18.6	989	4	AB60296 Drosophil
28	287	18.4	423	5	ABP43750 Ligase TT
29	287	18.4	423	8	ADJ93458 Human tub
30	281	18.0	673	8	ADQ66481 Novel hum
31	278	17.8	321	7	ADC33066 Human nov
32	278	17.8	393	7	ADC31411 Human nov
33	274	17.6	49	8	ADJ93373 Human BGS
34	272	17.4	487	4	AB59096 Drosophil
35	271.5	17.4	519	4	AAM79305 Human pro
36	271.5	17.4	524	4	ABB22915 Protein #
37	271.5	17.4	524	7	ABB64837 Human pro
38	270	17.3	917	4	ABB64837 Drosophil
39	256.5	16.4	439	7	ADB65504 Human pro
40	255.5	16.4	379	8	ADJ93456 Pig tubul
41	250.5	16.1	377	4	AAM79068 Human pro
42	250.5	16.1	377	5	ABB83472 Human cyt
43	250.5	16.1	377	7	ADE47756 Human NOV
44	250.5	16.1	377	8	ADJ79026 Human NOV
45	249.5	16.0	1138	8	ADJ71954 Human PMM

ALIGNMENTS

RESULT 1

ADJ93366	ID	ADJ93366	standard; protein; 293 AA.
XX	AC	ADJ93366;	
XX	DT	06-MAY-2004	(first entry)
XX	DE	Human BGS-42	protein-related TTL1 domain.
XX	KW	testis-specific tubulin tyrosine-ligase-like polypeptide;	
XX	KW	BGS-42 polypeptide; cytosolic; respiratory-Gen; gastrointestinal-Gen;	
XX	KW	neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;	
XX	KW	osteopathic; nootropic; antiparkinsonian; antiarthritic; antiasthmatic;	
XX	KW	anti-HIV; antibacterial; immunosuppressive; antiseborrheic;	
XX	KW	dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;	
XX	KW	tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;	
XX	KW	testicular disorder; testicular cancer; pulmonary disorder; lung cancer;	
XX	KW	gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;	
XX	KW	brain cancer; liver cancer; proliferative condition; testis; lung;	
XX	KW	small intestine; brain; lymph tissue; infertility; Cushing's syndrome;	
XX	KW	emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;	
XX	KW	Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;	
XX	KW	sepsis; acne; Sjogren's disease; scleroderma; human; TTL1 domain.	
XX	OS	Homo sapiens.	
XX	FN	WO2004005487-A2.	
XX	PD	15-JAN-2004.	
XX	PP	09-JUL-2003; 2003WO-US021605.	
XX	PP	09-JUL-2002; 2002US-0394725P.	
XX	PR	(BRIM) BRISTOL-MYERS SQUIBB CO.	
XX	PI	Feder JN, Wu S, Nelson TC;	
XX	PI	WPI; 2004-099381/10.	
XX	DR	New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,	
XX	PT	useful for preventing, treating or ameliorating a medical condition, e.g.	
XX	PT	aberrant cellular proliferation, reproductive disorders or testicular	
XX	PT	disorders.	
XX	PS	Disclosure; SEQ ID NO 14; 343pp; English.	

XX CC This invention relates to a novel testis-specific tubulin tyrosine-ligase
 CC -like polypeptide, designated the BGS-42 polypeptide. The invention may
 CC be useful for the development of compounds with a cytostatic, respiratory
 CC -Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,
 CC antiinflammatory, anabolic, hypertensive, osteopathic, nootropic,
 CC antiparkinsonian, antiarthritic, antiasthmatic, anti-HIV, antibacterial,
 CC immunosuppressive, antiseborrheic or dermatological activity acting as
 CC tyrosine ligase modulators. In addition, the disclosed sequences may be
 CC useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be
 CC used for diagnosing a pathological condition or a susceptibility to a
 CC pathological condition in a subject, and for preventing, treating or
 CC ameliorating a medical condition, such as a disorder related to aberrant tubulin-
 CC tubulin ligase activity, a disorder related to aberrant tubulin-
 CC carboxypeptidase activity, aberrant cellular proliferation, reproductive
 CC disorders, testicular disorders, testicular cancer, pulmonary disorders,
 CC lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
 CC neural disorders, brain cancer, liver cancer, or proliferative condition
 CC of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
 CC polypeptide, polynucleotide, or their modulators are also useful for
 CC treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
 CC disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
 CC -42 polypeptide can be used as a preventive agent for immunological
 CC disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
 CC disease or scleroderma. The antibodies may be used to purify, detect and
 CC target the BGS-42 polypeptides. The present sequence is that of the TTL1
 CC domain of the human BGS-42 protein of the invention.

XX SQ Sequence 293 AA;

Query Match 100.0%; Score 1560; DB 8; Length 293;
 Best Local Similarity 100.0%; Pred. No. 2.8e-160;
 Matches 293; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDIDTSADAVEDTEAEWEDLTQOYVSLVHGDAFISNRYFSQCQALLNRITSVNPQTD 60
 DB 1 EDIDTSADAVEDTEAEWEDLTQOYVSLVHGDAFISNRYFSQCQALLNRITSVNPQTD 60

QY 61 IDGLRNIIWIKPAKSRGRDIVCMVDRVVEILEAAADHPLSRDNKWKVQKIETPLLICD 120
 DB 61 IDGLRNIIWIKPAKSRGRDIVCMVDRVVEILEAAADHPLSRDNKWKVQKIETPLLICD 120

QY 121 TKFDIRQWFLVTDWNPNTIWFYKESYLRFSTQFSLDKLDSAIHLCCNNAVKYLNKDVGR 180
 DB 121 TKFDIRQWFLVTDWNPNTIWFYKESYLRFSTQFSLDKLDSAIHLCCNNAVKYLNKDVGR 180

QY 181 SPLLPAHNMWTSRFBQYLQQRGAVGWSVYFPMKKATAHAMKVAQDHVPERKNSFEL 240
 DB 181 SPLLPAHNMWTSRFBQYLQQRGAVGWSVYFPMKKATAHAMKVAQDHVPERKNSFEL 240

QY 241 YGADFVLGRDRFRLWLEINSPTHPSTPTVTAQLCAQVQEDTIKVAVDKSCDI 293
 DB 241 YGADFVLGRDRFRLWLEINSPTHPSTPTVTAQLCAQVQEDTIKVAVDKSCDI 293

RESULT 2

ADJ93365

ID ADJ93365 standard; protein; 541 AA.

AC ADJ93365;

XX 06-MAY-2004 (first entry)

XX Human tubulin tyrosine ligase protein consensus sequence SeqID13.

XX testis-specific tubulin tyrosine-ligase-like polypeptide;
 KW BGS-42 polypeptide; cytosolic; respiratory-Gen; gastrointestinal-Gen;
 KW neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;
 KW osteopathic; nootropic; antiparkinsonian; antiarthritic; antiasthmatic;
 KW anti-HIV; antibacterial; immunosuppressive; antiseborrheic;
 KW dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;
 KW tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;
 KW testicular disorder; testicular cancer; pulmonary disorder; lung cancer;

KW gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;
 KW brain cancer; liver cancer; proliferative condition; testis; lung;
 KW small intestine; brain; lymph tissue; infertility; Cushing's syndrome;
 KW emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;
 KW Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;
 KW sepsis; acne; Sjogren's disease; scleroderma; human.

XX Homo sapiens.
 OS Synthetic.
 XX WO2004005487-A2.
 XX 15-JAN-2004.
 PD 09-JUL-2003; 2003WO-US021605.
 PF 09-JUL-2002; 2002US-0394725P.
 PR (BRIM) BRISTOL-MYERS SQUIBB CO.
 PA Feder JN, Wu S, Nelson TC;
 PI WPI; 2004-099381/10.
 DR N-PSDB; ADJ93364.
 XX New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,
 PT useful for preventing, treating or ameliorating a medical condition, e.g.
 PT aberrant cellular proliferation, reproductive disorders or testicular
 XX disorders.

XX Example 4; SEQ ID NO 13; 343pp; English.

XX This invention relates to a novel testis-specific tubulin tyrosine-ligase
 CC -like polypeptide, designated the BGS-42 polypeptide. The invention may
 CC be useful for the development of compounds with a cytostatic, respiratory
 CC -Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,
 CC antiinflammatory, anabolic, hypertensive, osteopathic, nootropic,
 CC antiparkinsonian, antiarthritic, antiasthmatic, anti-HIV, antibacterial,
 CC immunosuppressive, antiseborrheic or dermatological activity acting as
 CC tyrosine ligase modulators. In addition, the disclosed sequences may be
 CC useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be
 CC used for diagnosing a pathological condition or a susceptibility to a
 CC pathological condition in a subject, and for preventing, treating or
 CC ameliorating a medical condition, such as a disorder related to aberrant tubulin-
 CC tubulin ligase activity, a disorder related to aberrant tubulin-
 CC carboxypeptidase activity, aberrant cellular proliferation, reproductive
 CC disorders, testicular disorders, testicular cancer, pulmonary disorders,
 CC lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
 CC neural disorders, brain cancer, liver cancer, or proliferative condition
 CC of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
 CC polypeptide, polynucleotide, or their modulators are also useful for
 CC treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
 CC disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
 CC -42 polypeptide can be used as a preventive agent for immunological
 CC disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
 CC disease or scleroderma. The antibodies may be used to purify, detect and
 CC target the BGS-42 polypeptides. The present sequence is that of the
 CC tubulin tyrosine ligase protein consensus sequence which was used in the
 CC exemplification of the invention.

XX SQ Sequence 541 AA;

Query Match 100.0%; Score 1560; DB 8; Length 541;
 Best Local Similarity 100.0%; Pred. No. 7e-160;
 Matches 293; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDIDTSADAVEDTEAEWEDLTQOYVSLVHGDAFISNRYFSQCQALLNRITSVNPQTD 60
 DB 73 EDIDTSADAVEDTEAEWEDLTQOYVSLVHGDAFISNRYFSQCQALLNRITSVNPQTD 132

QY 61 IDGLRNIIWIKPAKSRGRDIVCMVDRVVEILEAAADHPLSRDNKWKVQKIETPLLICD 120
 DB 133 IDGLRNIIWIKPAKSRGRDIVCMVDRVVEILEAAADHPLSRDNKWKVQKIETPLLICD 192

QY 121 TKFDIRQWFLVTDWNPDLTIWFYKESYLRFSTQSFSLDKLDSAIHLGNNAVKYLNKDVGR 180
 DB 193 TKFDIRQWFLVTDWNPDLTIWFYKESYLRFSTQSFSLDKLDSAIHLGNNAVKYLNKDVGR 252
 QY 181 SPLLPAHNMTSTRFQEYLQROGRGAVGWSVIYPSMKATAHAKVQAQDHVPRKNSFEL 240
 DB 253 SPLLPAHNMTSTRFQEYLQROGRGAVGWSVIYPSMKATAHAKVQAQDHVPRKNSFEL 312
 QY 241 YGADFVLGRDRPWLIEINSSPTMHPSTPTVTAQLCAQVQEDTIKVAVDRSCDI 293
 DB 313 YGADFVLGRDRPWLIEINSSPTMHPSTPTVTAQLCAQVQEDTIKVAVDRSCDI 365

RESULT 3

ADJ93358
 ID ADJ93358 standard; protein; 541 AA.

XX AC ADJ93358;

XX DT 06-MAY-2004 (first entry)

XX DE Human BGS-42 protein sequence SeqID2.

XX testis-specific tubulin tyrosine-ligase-like polypeptide;
 KW BGS-42 polypeptide; cytosolic; respiratory-Gen; gastrointestinal-Gen;
 KW neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;
 KW osteopathic; nootropic; antiparkinsonian; antiarthritic; antiasthmatic;
 KW anti-HIV; antibacterial; immunosuppressive; antiseborrheic;
 KW dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;
 KW tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;
 KW testicular disorder; testicular cancer; pulmonary disorder; lung cancer;
 KW gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;
 KW brain cancer; liver cancer; proliferative condition; testis; lung;
 KW small intestine; brain; lymph tissue; infertility; Cushing's syndrome;
 KW emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;
 KW Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;
 KW sepsis; acne; Sjogren's disease; scleroderma; human.

OS Homo sapiens.

XX WO2004005487-A2.

XX 15-JAN-2004.

XX 09-JUL-2003; 2003WO-US021605.

XX 09-JUL-2002; 2002US-0394725P.

XX (BRIM) BRISTOL-MYERS SQUIBB CO.

XX Feder JN, Wu S, Nelson TC;

XX WPI: 2004-099381/10.

DR N-PSDB; ADJ93357.

XX New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,
 PT useful for preventing, treating or ameliorating a medical condition, e.g.
 PT aberrant cellular proliferation, reproductive disorders or testicular
 PT disorders.

XX Claim 5; SEQ ID NO 2; 343pp; English.

XX This invention relates to a novel testis-specific tubulin tyrosine-ligase
 CC -like polypeptide, designated the BGS-42 polypeptide. The invention may
 CC be useful for the development of compounds with a cytosolic, respiratory
 CC -Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,
 CC antiinflammatory, anabolic, hypertensive, osteopathic, nootropic,
 CC antiparkinsonian, antiarthritic, antiasthmatic, anti-HIV, antibacterial,
 CC immunosuppressive, antiseborrheic or dermatological activity acting as
 CC tyrosine ligase modulators. In addition, the disclosed sequences may be
 CC useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be
 CC used for diagnosing a pathological condition or a susceptibility to a

CC pathological condition in a subject, and for preventing, treating or
 CC ameliorating a medical condition, such as a disorder related to aberrant
 CC tubulin ligase activity, a disorder related to aberrant tubulin-
 CC carboxypeptidase activity, aberrant cellular proliferation, reproductive
 CC disorders, testicular disorders, testicular cancer, pulmonary disorders,
 CC lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
 CC neural disorders, brain cancer, liver cancer, or proliferative condition
 CC of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
 CC polypeptide, polynucleotide, or their modulators are also useful for
 CC treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
 CC disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
 CC -42 polypeptide can be used as a preventive agent for immunological
 CC disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
 CC disease or scleroderma. The antibodies may be used to purify, detect and
 CC target the BGS-42 polypeptides. The present sequence is that of the human
 CC BGS-42 protein of the invention.

XX SQ Sequence 541 AA;

Query Match 100.0%; Score 1560; DB 8; Length 541;
 Best Local Similarity 100.0%; Pred. No. 7e-160;
 Matches 293; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDIDTSADAVEDLTEAEWEDLTQQYYSLVHGDAFISNSRNYFSQCQALLNRITSVNPQTD 60
 DB 73 EDIDTSADAVEDLTEAEWEDLTQQYYSLVHGDAFISNSRNYFSQCQALLNRITSVNPQTD 132
 QY 61 IDGLRNWITIKPAKSRGRDIVCMRVEREILAAADHPLSRDNKVVVKYIETPLLICD 120
 DB 133 IDGLRNWITIKPAKSRGRDIVCMRVEREILAAADHPLSRDNKVVVKYIETPLLICD 192
 QY 121 TKFDIRQWFLVTDWNPDLTIWFYKESYLRFSTQSFSLDKLDSAIHLGNNAVKYLNKDVGR 180
 DB 193 TKFDIRQWFLVTDWNPDLTIWFYKESYLRFSTQSFSLDKLDSAIHLGNNAVKYLNKDVGR 252
 QY 181 SPLLPAHNMTSTRFQEYLQROGRGAVGWSVIYPSMKATAHAKVQAQDHVPRKNSFEL 240
 DB 253 SPLLPAHNMTSTRFQEYLQROGRGAVGWSVIYPSMKATAHAKVQAQDHVPRKNSFEL 312
 QY 241 YGADFVLGRDRPWLIEINSSPTMHPSTPTVTAQLCAQVQEDTIKVAVDRSCDI 293
 DB 313 YGADFVLGRDRPWLIEINSSPTMHPSTPTVTAQLCAQVQEDTIKVAVDRSCDI 365

RESULT 4

ADJ93360
 ID ADJ93360 standard; protein; 242 AA.

XX AC ADJ93360;

XX DT 06-MAY-2004 (first entry)

XX DE Human BGS-42 protein sequence SeqID4.

XX testis-specific tubulin tyrosine-ligase-like polypeptide;
 KW BGS-42 polypeptide; cytosolic; respiratory-Gen; gastrointestinal-Gen;
 KW neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;
 KW osteopathic; nootropic; antiparkinsonian; antiarthritic; antiasthmatic;
 KW anti-HIV; antibacterial; immunosuppressive; antiseborrheic;
 KW dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;
 KW tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;
 KW testicular disorder; testicular cancer; pulmonary disorder; lung cancer;
 KW gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;
 KW brain cancer; liver cancer; proliferative condition; testis; lung;
 KW small intestine; brain; lymph tissue; infertility; Cushing's syndrome;
 KW emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;
 KW Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;
 KW sepsis; acne; Sjogren's disease; scleroderma; human.

OS Homo sapiens.

XX WO2004005487-A2.

XX

PD 15-JAN-2004.
XX
XX
PF 09-JUL-2003; 2003WO-US021605.
XX
XX PR 09-JUL-2002; 2002US-0394725P.
XX
XX (BRIM) BRISTOL-MYERS SQUIBB CO.
PA
PA Feder JN, Wu S, Nelson TC;
PI
XX WPI; 2004-099381/10.
XX
XX New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,
PT useful for preventing, treating or ameliorating a medical condition, e.g.
PT aberrant cellular proliferation, reproductive disorders or testicular
PT disorders.
XX
XX PS Disclosure; SEQ ID NO 4; 343pp; English.
XX
XX This invention relates to a novel testis-specific tubulin tyrosine-ligase
CC -like polypeptide, designated the BGS-42 polypeptide. The invention may
CC be useful for the development of compounds with a cytostatic, respiratory
CC -Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,
CC antiinflammatory, anabolic, hypertensive, osteopathic, nootropic,
CC antiparkinsonian, antiarthritic, antiaesthetic, anti-HIV, antibacterial,
CC immunosuppressive, antiseborrheic or dermatological activity acting as
CC tyrosine ligase modulators. In addition, the disclosed sequences may be
CC useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be
CC used for diagnosing a pathological condition or a susceptibility to a
CC pathological condition in a subject, and for preventing, treating or
CC ameliorating a medical condition, such as a disorder related to aberrant
CC tubulin ligase activity, a disorder related to aberrant tubulin-
CC carboxypeptidase activity, aberrant cellular proliferation, reproductive
CC disorders, testicular disorders, testicular cancer, pulmonary disorders,
CC lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
CC neural disorders, brain cancer, liver cancer, or proliferative condition
CC of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
CC polypeptide, polynucleotide, or their modulators are also useful for
CC treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
CC disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
CC -42 polypeptide can be used as a preventive agent for immunological
CC disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
CC disease or scleroderma. The antibodies may be used to purify, detect and
CC target the BGS-42 polypeptides. The present sequence is that of the human
XX BGS-42 protein (partial sequence) of the invention.
XX
SQ Sequence 242 AA;

Query Match 79.9%; Score 1247; DB 8; Length 242;
Best Local Similarity 100.0%; Pred. No. 2.2e-126; Indels 0; Gaps 0;
Matches 233; Conservative 0; Mismatches 0;

QY 61 IDGLRNITWIKPAKSRGRDIVCMRVBEIELEAAADHPLSRDNKVVQVYIETPLICD 120
Db 1 IDGLRNITWIKPAKSRGRDIVCMRVBEIELEAAADHPLSRDNKVVQVYIETPLICD 60
QY 121 TKFDIRQWFLVTDWNPITWIFKESYLRFSTQPSLKLDSAIHLNNVAKYLNVDVGR 180
Db 61 TKFDIRQWFLVTDWNPITWIFKESYLRFSTQPSLKLDSAIHLNNVAKYLNVDVGR 120
QY 181 SPLLPANHMTSTFOVLQGRGAVWGSVIVPSMKKATAHMKVAQDHVEPRKNSFEL 240
Db 121 SPLLPANHMTSTFOVLQGRGAVWGSVIVPSMKKATAHMKVAQDHVEPRKNSFEL 180
QY 241 YGADFVLGRDPRPWLIEINSPTMHPSTPVTQAQCAQVQEDTIKVAVDNRSCDI 293
Db 181 YGADFVLGRDPRPWLIEINSPTMHPSTPVTQAQCAQVQEDTIKVAVDNRSCDI 233

RESULT 5

AAB43005

ID AAB43005 standard; protein; 352 AA.

XX

AC AAB43005;
XX
XX DT 08-FEB-2001 (first entry)
XX
XX DE Human ORFX ORF2769 polypeptide sequence SEQ ID NO:5538.
XX
XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnery; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antinflammatory disease; coagulation;
KW thrombosis; contraceptive.
XX
XX Homo sapiens.
OS
XX WO200058473-A2.
XX
XX PD 05-OCT-2000.
XX
XX PF 31-MAR-2000; 2000WO-US008621.
XX
XX PR 31-MAR-1999; 99US-0127607P.
XX PR 02-APR-1999; 99US-0127636P.
XX PR 05-APR-1999; 99US-0127728P.
XX PR 30-MAR-2000; 2000US-00540763.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Shimkets RA, Leach M;
PI WPI; 2000-602362/57.
DR N-PSDB; AAC77214.
XX
XX Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease.
XX
XX Claim 11; Page 4720-4721; 5507pp; English.
XX
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective; osteopathic;
CC anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;
CC cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
CC dermatological; immunosuppressive; antinflammatory; antibacterial;
CC antiviral; antifungal; antirheumatic; antithyroid; and antianaemic. The
CC sequences can be used for determining the presence of or predisposition
CC to, or preventing or treating pathological conditions associated with an
CC ORFX-associated disorder. The nucleic acids can be used to express ORFX
CC proteins in gene therapy vectors. The proteins and nucleic acids may be
CC used to treat cancers, proliferative disorders, neurodegenerative
CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
CC storage, systemic lupus erythematosus, severe combined immunodeficiency
CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
CC cartilage damage, nocturnal haemoglobinuria, antinflammatory disease; to
CC enhance coagulation; to inhibit thrombosis; and as a contraceptive
XX
XX SQ Sequence 352 AA;

Query Match 58.2%; Score 908; DB 3; Length 352;
Best Local Similarity 56.7%; Pred. No. 2.6e-89;
Matches 165; Conservative 48; Mismatches 78; Indels 0; Gaps 0;

QY 2 DIDSADAVEDL TEAEWEDLTQYYSLVHGDAFISNRNYFSQCQALLNRITSVNPQTDI 61
 DB 2 DIDKOLEAPLYLTPEGNSLFQYYQVHGEAELRHLDTQVQRCEDILQOLQAVVPQIDM 61
 QY 62 DGLRNIWIIPKAASGRDIVCMQDRVEEILEAAADHPLSRDKKVVQKIETPLLCIDT 121
 DB 62 EGRNIWIIVKPAKSRGRGIMCMHLEMLKLVNGVPVVKDGKVVQKIETPLIFGT 121
 QY 122 KFDIRQWFLVTDWNPPLTIWFKESYLRFSTORFSLDKLDSAIHLGNNAVQKYLKNDVGRS 181
 DB 122 KFDLRQWFLVTDWNPPLTIWFKESYLRFSTORFSLDKLDSAIHLGNNSIQKLENSCHRH 181
 QY 182 PLLPHNNMTSTRFOEYLQROGRGAVGWSVIYPSMKKAIAHAMKVAQDHVEPRKNSFELY 241
 DB 182 PLLPPDNWSSORFOAHLQEMGAPNAWSTIIIVGKMDAVIHALQTSQDTVQCRKASFELY 241
 QY 242 GADFVGLGRDPRFWLIEINSSPTMHPSTPTVTAOLCAQVOEDTIKVAVDKSCD 292
 DB 242 GADFVFGDFQFWLIEINASPTMAPSTAVTARLCAGVQADTLRVVIDRRLD 292

RESULT 6
 AAM39450
 ID AAM39450 standard; protein; 352 AA.
 AC AAM39450;
 XX
 DT 22-OCT-2001 (first entry)
 DE Human polypeptide SEQ ID NO 2595.
 XX
 KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.
 XX
 OS Homo sapiens.
 XX
 PN WO200153312-A1.
 XX
 PD 26-JUL-2001.
 XX
 PF 26-DEC-2000; 2000WO-US034263.
 XX
 PR 23-DEC-1999; 99US-00471275.
 PR 21-JAN-2000; 2000US-00488725.
 PR 25-APR-2000; 2000US-0052317.
 PR 20-JUN-2000; 2000US-00598042.
 PR 19-JUL-2000; 2000US-00620312.
 PR 03-AUG-2000; 2000US-00653450.
 PR 14-SEP-2000; 2000US-00662191.
 PR 19-OCT-2000; 2000US-00693036.
 PR 29-NOV-2000; 2000US-00727344.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, xue AJ, Yang Y, Zhang J, Zhao QA;
 PI Zhou P, Goodrich R, Drmanac RT;
 XX
 XX WPI; 2001-442253/47.
 DR N-PSDB; AAI58606.
 XX
 XX Novel nucleic acids and polypeptides, useful for treating disorders such
 FT as central nervous system injuries.
 XX
 XX Example 4; SEQ ID NO 2595; 10078pp; English.
 PS
 CC The invention relates to human nucleic acids (AAI57798-AAI61369) and the
 CC encoded polypeptides (AAM39642-AA42213) with nootropic,

CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localized neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening
 CC assays for receptor activity, arthritis and inflammation, leukaemia and
 CC C.N.S disorders. Note: The sequence data for this patent did not form
 CC part of the printed specification

XX Sequence 352 AA;

Query Match 58.2%; Score 908; DB 4; Length 352;
 Best Local Similarity 56.7%; Pred. No. 2.6e-89;
 Matches 165; Conservative 48; Mismatches 78; Indels 0; Gaps 0;

QY 2 DIDSADAVEDL TEAEWEDLTQYYSLVHGDAFISNRNYFSQCQALLNRITSVNPQTDI 61
 DB 2 DIDKOLEAPLYLTPEGNSLFQYYQVHGEAELRHLDTQVQRCEDILQOLQAVVPQIDM 61
 QY 62 DGLRNIWIIPKAASGRDIVCMQDRVEEILEAAADHPLSRDKKVVQKIETPLLCIDT 121
 DB 62 EGRNIWIIVKPAKSRGRGIMCMHLEMLKLVNGVPVVKDGKVVQKIETPLIFGT 121
 QY 122 KFDIRQWFLVTDWNPPLTIWFKESYLRFSTORFSLDKLDSAIHLGNNAVQKYLKNDVGRS 181
 DB 122 KFDLRQWFLVTDWNPPLTIWFKESYLRFSTORFSLDKLDSAIHLGNNSIQKLENSCHRH 181
 QY 182 PLLPHNNMTSTRFOEYLQROGRGAVGWSVIYPSMKKAIAHAMKVAQDHVEPRKNSFELY 241
 DB 182 PLLPPDNWSSORFOAHLQEMGAPNAWSTIIIVGKMDAVIHALQTSQDTVQCRKASFELY 241
 QY 242 GADFVGLGRDPRFWLIEINSSPTMHPSTPTVTAOLCAQVOEDTIKVAVDKSCD 292
 DB 242 GADFVFGDFQFWLIEINASPTMAPSTAVTARLCAGVQADTLRVVIDRRLD 292

RESULT 7

ADJ93457
 ID ADJ93457 standard; protein; 352 AA.
 AC ADJ93457;
 XX
 DT 06-MAY-2004 (first entry)
 DE Human HOTTLL3 protein sequence SeqID2.
 XX
 KW testis-specific tubulin tyrosine-ligase-like polypeptide;
 KW BGS-42 polypeptide; cytostatic; respiratory-Gen; gastrointestinal-Gen;
 KW neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;
 KW osteopathic; nootropic; antiparkinsonian; antiarthritic; antiasthmatic;
 KW anti-HIV; antibacterial; immunosuppressive; antiseborrheic;
 KW dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;
 KW tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;
 KW testicular disorder; testicular cancer; pulmonary disorder; lung cancer;
 KW gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;
 KW brain cancer; liver cancer; proliferative condition; testis; lung;
 KW small intestine; brain; lymph tissue; infertility; Cushing's syndrome;
 KW emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;
 KW Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;
 KW sepsis; acne; Sjogren's disease; scleroderma; human; HOTTLL3.
 XX
 OS Homo sapiens.
 XX
 PN WO2004005487-A2.
 XX
 PD 15-JAN-2004.
 XX
 PF 09-JUL-2003; 2003WO-US021605.

XX PR 09-JUL-2002; 2002US-0394725P.
XX PA (BRIM) BRISTOL-MYERS SQUIBB CO.
XX PI Feder JN, Wu S, Nelson TC;
XX WPI; 2004-099381/10.
XX
XX New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,
PT useful for preventing, treating or ameliorating a medical condition, e.g.
PT aberrant cellular proliferation, reproductive disorders or testicular
PT disorders.
XX
XX Disclosure; SEQ ID NO 7; 343pp; English.
XX
XX This invention relates to a novel testis-specific tubulin tyrosine-ligase
CC -like polypeptide, designated the BGS-42 polypeptide. The invention may
CC be useful for the development of compounds with a cytostatic, respiratory
CC -Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,
CC antiinflammatory, anabolic, hypertensive, osteopathic, nootropic,
CC antiparkinsonian, antiarthritic, antiasthmatic, anti-HIV, antibacterial,
CC immunosuppressive, antiseborrheic or dermatological activity acting as
CC tyrosine ligase modulators. In addition, the disclosed sequences may be
CC useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be
CC used for diagnosing a pathological condition or a susceptibility to a
CC pathological condition in a subject, and for preventing, treating or
CC ameliorating a medical condition, such as a disorder related to aberrant
CC tubulin ligase activity, a disorder related to aberrant tubulin-
CC carboxypeptidase activity, aberrant cellular proliferation, reproductive
CC disorders, testicular disorders, testicular cancer, pulmonary disorders,
CC lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
CC neural disorders, breast cancer, liver cancer, or proliferative condition
CC of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
CC polypeptide, polynucleotide, or their modulators are also useful for
CC treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
CC disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
CC -42 polypeptide can be used as a preventive agent for immunological
CC disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
CC disease or scleroderma. The antibodies may be used to purify, detect and
CC target the BGS-42 polypeptides. The present sequence is that of the human
CC HOTTU3 protein which is related to the invention. Note: The present
CC sequence does not appear in the specification but was obtained from
CC Genbank.
XX
XX Sequence 352 AA;

Query Match 58.2%; Score 908; DB 8; Length 352;
Best Local Similarity 56.7%; Pred. No. 2.6e-89;
Matches 165; Conservative 48; Mismatches 78; Indels 0; Gaps 0;

QY 2 DITSDADVEDLTEAEWEDLTQQYSLVHGDAFISNRNYSFQCQALLNRITSVNPQTDI 61
DB 2 DIDKDLEAPLYLTPEGWSLFQRYQVVGHEGAEIRHLDTQVQRCELIQQLQAVVQIDM 61
QY 62 DGLRNIWIIPAAKSRGRDIVCMRVEIELEAAADHPLSRDNKRWVQVYETPLLICDT 121
DB 62 EGDRIWIIVKPGAKSRGIMCMCHLEEMKLVNGNPVVMKDGKVVQVQYIERPLIFGT 121
QY 122 KFDIRQWFLTDNPLTIWYKSYLRFSTQSRSLDKLSAHLGNNAVOKYLNVDGRS 181
DB 122 KFDLRQWFLTDNPLTVWFYRDSYIRFSTQPSFLKNLSDNSVHLNNSIQKHLNSENCHRH 181
QY 182 PLLPAHNMWTSRFOEYLQGRGAVGVMSVTPYSMKKAIAMKVAQDHVEPRKNSPELY 241
DB 182 PLLPPDNNWSQRQFALHQEWGAPNASTIIVFGMKDAVHALQTSQDTVQCRKASPELY 241
QY 242 GADPVLGRDFRPMLEINSSPTHPSTFVTAQLCAQOQEDTIKAVYDRSDC 292
DB 242 GADPVFGEEDQFPMLEINASPTMAPSTAVTARLCAGVQADTLRVVIDRMLD 292

ABM80420
ID ABM80420 standard; protein; 352 AA.
XX
AC ABM80420;
XX
DT 18-NOV-2004 (first entry)
XX
DE Tumour-associated antigenic target (TAT) polypeptide PRO80956, SEQ:1058.
XX
KW Tumour-associated antigenic target; TAT; human; overexpression; cancer;
KW tumour; diagnosis; cell proliferative disorder; breast cancer;
KW colorectal cancer; lung cancer; ovarian cancer; liver cancer;
KW central nervous system cancer; bladder cancer; pancreatic cancer;
KW cervical cancer; melanoma; leukaemia; hybridisation probe;
KW chromosome identification; chromosome mapping; gene mapping;
KW gene therapy; cytostatic.
XX
OS Homo sapiens.
XX
PN WO2004030615-A2.
XX
PD 15-APR-2004.
XX
PF 29-SEP-2003; 2003WO-US028547.
XX
PR 02-OCT-2002; 2002US-0414971P.
XX
PA (GETH) GENENTECH INC.
XX
PI Wu TD, Zhang Z, Zhou Y;
XX
WPI; 2004-347921/32.
XX
N-PSDB; ACN37881.
XX
New tumor-associated antigenic target polypeptides and nucleic acids,
PT useful in preparing a medicament for treating or detecting a
PT proliferative disorder, e.g. breast, lung, colorectal, ovarian or
PT prostate cancer or tumor.
XX
PS Claim 12; SEQ ID NO 1058; 7273pp; English.
XX
XX The invention relates to human tumour-associated antigenic target (TAT)
CC polypeptides, and their related nucleic acids. The TAT polypeptides are
CC overexpressed in cancer tissues compared to normal tissues, and may thus
CC serve as effective targets for the diagnosis and treatment of cancer in
CC mammals. The invention also relates to nucleic acid and polypeptide
CC sequences at least 80% identical to the TAT nucleic acids and
CC polypeptides; expression vectors and host cells comprising a TAT nucleic
CC acid; an antibody specific for a TAT polypeptide; a peptide or organic
CC molecule which binds to a TAT polypeptide; fusion proteins comprising a
CC TAT polypeptide; and methods and compositions for the treatment or
CC diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,
CC antibodies, antagonists, binding molecules and host cells comprising a TAT
CC for diagnosing or treating a cell proliferative disorder associated with
CC increased TAT expression, particularly cancers such as breast cancer,
CC colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder
CC cancer, pancreatic cancer, cervical cancer, cancers of the central
CC nervous system, melanoma and leukaemia. TAT nucleic acids may further be
CC used as hybridisation probes, in chromosome and gene mapping, in
CC chromosome identification and in gene therapy. The present sequence
CC represents a TAT polypeptide of the invention
XX
XX Sequence 352 AA;

Query Match 58.2%; Score 908; DB 8; Length 352;
Best Local Similarity 56.7%; Pred. No. 2.6e-89;
Matches 165; Conservative 48; Mismatches 78; Indels 0; Gaps 0;

QY 2 DITSDADVEDLTEAEWEDLTQQYSLVHGDAFISNRNYSFQCQALLNRITSVNPQTDI 61
DB 2 DIDKDLEAPLYLTPEGWSLFQRYQVVGHEGAEIRHLDTQVQRCELIQQLQAVVQIDM 61
QY 62 DGLRNIWIIPAAKSRGRDIVCMRVEIELEAAADHPLSRDNKRWVQVYETPLLICDT 121
DB 62 EGDRIWIIVKPGAKSRGIMCMCHLEEMKLVNGNPVVMKDGKVVQVQYIERPLIFGT 121
QY 122 KFDIRQWFLTDNPLTIWYKSYLRFSTQSRSLDKLSAHLGNNAVOKYLNVDGRS 181
DB 122 KFDLRQWFLTDNPLTVWFYRDSYIRFSTQPSFLKNLSDNSVHLNNSIQKHLNSENCHRH 181
QY 182 PLLPAHNMWTSRFOEYLQGRGAVGVMSVTPYSMKKAIAMKVAQDHVEPRKNSPELY 241
DB 182 PLLPPDNNWSQRQFALHQEWGAPNASTIIVFGMKDAVHALQTSQDTVQCRKASPELY 241
QY 242 GADPVLGRDFRPMLEINSSPTHPSTFVTAQLCAQOQEDTIKAVYDRSDC 292
DB 242 GADPVFGEEDQFPMLEINASPTMAPSTAVTARLCAGVQADTLRVVIDRMLD 292

```

Db 62 EGRNIIWIKPAKSGRGLMCMHLEMLKLVGNPVMKDGKVVQYKIERPLIFGT 121
QY 122 KFDIRQWFLVTDWNPITWYFYESYLFSTQSFSLDKLDSAIHLCCNNAVQKYLKNDVGRS 181
Db 122 KFDLRQWFLVTDWNPITWYFYESYLFSTQSFSLDKLDSAIHLCCNNAVQKYLKNDVGRS 181
QY 182 PLLPAHNNWTSTRFOEYLQROGRGAVGWSVIYPSMKKAIAHAMKVAQDHVEPRKNSFELY 241
Db 182 PLLPPDNMWSQRFOAHLQEMGAPNAWSTIIIVFGMKDAVIHALQTSQDTVQCRKASFELY 241
QY 242 GADFVLGRDRFPLWLIENSSPTMHPSTVTAQIACAQVEDTIKVAVDKSCD 292
Db 242 GADFVGEDFQFWLIEINASTMAPSTAVTARLCAGVQADTLKRVVDRMLD 292

RESULT 9
ABU11512
ID ABU11512 standard; protein; 399 AA.
XX AC ABU11512;
XX DT 12-FEB-2003 (first entry)
XX DE Human MDDT polypeptide SEQ ID 459.
XX KW MDDT; human; disease detection and treatment molecule polypeptide;
KW anti-inflammatory; immunosuppressive; osteopathic; cytostatic; anti-HIV;
KW haemostatic; nephrotropic; antianaemic; antipsoriatic; hepatotropic;
KW gene therapy; protein replacement therapy; cell proliferative disorder;
KW cancer; adenocarcinoma; leukaemia; lymphoma; melanoma; myeloma; sarcoma;
KW anaemia; Crohn's disease; acquired immunodeficiency syndrome; AIDS;
KW Goodpasture's syndrome; inflammation; osteoporosis; thrombocytopaenia;
KW psoriasis; hepatitis.
XX OS Homo sapiens.
XX PN WO200279449-A2.
XX PD 10-OCT-2002.
XX PF 27-MAR-2002; 2002WO-US009944.
XX PR 28-MAR-2001; 2001US-0279619P.
XX PR 29-MAR-2001; 2001US-0280067P.
XX PR 29-MAR-2001; 2001US-0280068P.
XX PR 16-MAY-2001; 2001US-0291280P.
XX PR 17-MAY-2001; 2001US-0291829P.
XX PR 17-MAY-2001; 2001US-0291849P.
XX PR 19-JUN-2001; 2001US-0299428P.
XX PR 20-JUN-2001; 2001US-0299776P.
XX PR 20-JUN-2001; 2001US-030001P.
XX PA (INCY-) INCYTE GENOMICS INC.
XX PI Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J,
PI Dufour GE, Hillman JL, Yu JY, Tuason O, Yap PE, Amshey SR,
PI Daugherty SC, Dam TC, Liu TF, Nguyen DA, Kleefeld Y, Gerstin EH,
PI Peralta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B,
PI Flores V, Marwaha R, Lo A, Lan RY, Urashka ME;
XX WPI; 2003-058431/05.
XX DR N-PSDB; ABX34502.
XX PT New purified disease detection and treatment molecule proteins and
XX PT polynucleotides, useful for diagnosing, treating or preventing cancers
XX PT (e.g. leukemia or sarcoma), anemia, Crohn's disease, AIDS, osteoporosis
XX PT or hepatitis.
XX PS Claim 27; SEQ ID NO 459; 339pp + Sequence Listing; English.
XX CC This invention describes a novel disease detection and treatment molecule
XX CC polypeptide (MDDT) which has anti-inflammatory, immunosuppressive,

```

```

CC osteopathic, cytostatic, anti-HIV, haemostatic, nephrotropic,
CC antianaemic, antipsoriatic and hepatotropic activity. The polynucleotides
CC and the polypeptides of the invention can be used for gene therapy, of
CC protein replacement therapy and are useful for treating a variety of
CC diseases or conditions. These polypeptides or polynucleotides are
CC particularly useful for diagnosing, treating or preventing cell
CC proliferative disorders (e.g. cancers including adenocarcinoma,
CC leukaemia, lymphoma, melanoma, myeloma or sarcoma), anaemia, Crohn's
CC disease, acquired immunodeficiency syndrome (AIDS), Goodpasture's
CC syndrome, inflammation, osteoporosis, thrombocytopaenia, psoriasis or
CC hepatitis. ABU11450-ABU11845 represent the MDDT polynucleotides encoded
CC by ABU11450-ABU11845, described in the disclosure of the invention. NOTE:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 399 AA;

```

```

Query Match 58.2%; Score 908; DB 6; Length 399;
Best Local Similarity 56.7%; Pred. No. 3.1e-89;
Matches 165; Conservative 48; Mismatches 78; Indels 0; Gaps 0;

QY 2 DDTSDADAVEDLTEAEWEDLTQOYYSLVHGDFAFISNRNYSFQCAALLNRITSVNPQTDI 61
Db 49 DIDKLEAPLYLTPEGWSLFLQRYQVHVEGAELRHLDTPQVQRCEDILQQLQAVVPQIDM 108

QY 62 DGLRNIWIKPAKSGRGLMCMHLEMLKLVGNPVMKDGKVVQYKIERPLIFGT 121
Db 109 EGRNIIWIKPAKSGRGLMCMHLEMLKLVGNPVMKDGKVVQYKIERPLIFGT 168

QY 122 KFDIRQWFLVTDWNPITWYFYESYLFSTQSFSLDKLDSAIHLCCNNAVQKYLKNDVGRS 181
Db 169 KFDLRQWFLVTDWNPITWYFYESYLFSTQSFSLDKLDSAIHLCCNNAVQKYLKNDVGRS 228

QY 182 PLLPAHNNWTSTRFOEYLQROGRGAVGWSVIYPSMKKAIAHAMKVAQDHVEPRKNSFELY 241
Db 229 PLLPPDNMWSQRFOAHLQEMGAPNAWSTIIIVFGMKDAVIHALQTSQDTVQCRKASFELY 288

QY 242 GADFVLGRDRFPLWLIENSSPTMHPSTVTAQIACAQVEDTIKVAVDKSCD 292
Db 289 GADFVGEDFQFWLIEINASTMAPSTAVTARLCAGVQADTLKRVVDRMLD 339

```

```

RESULT 10
AAU74334
ID AAU74334 standard; protein; 488 AA.
XX AC AAU74334;
XX DT 12-MAR-2002 (first entry)
XX DE Human cytoskeleton-associated protein (CYSKP) #5.
XX KW Human; cytoskeleton-associated protein; CYSKP; autoimmune disorder;
KW cell proliferative disorder; inflammatory disorder; prion disease;
KW vesicle trafficking disorder; gastrointestinal disorder; muscle disorder;
KW neurological disorder; cell motility disorder; reproductive disorder;
KW spinal cord disease; central nervous system disorder; mental disorder;
XX gene therapy; cancer.
XX OS Homo sapiens.
XX PN WO200185942-A2.
XX PD 15-NOV-2001.
XX PF 03-MAY-2001; 2001WO-US014355.
XX PR 05-MAY-2000; 2000US-0201960P.
XX PR 08-MAY-2000; 2000US-020279P.
XX PR 05-JUN-2000; 2000US-0209705P.
XX PR 07-JUN-2000; 2000US-0210149P.
XX PR 21-JUN-2000; 2000US-0213215P.

```

XX (INCY-) INCYTE GENOMICS INC.

XX Yue H, Tang YT, Au-Young J, Lu DAM, Baughn MR, Hillman JL;

PI Azimzai Y, Lal P, Yao MG, Bandman O, Burford N, Batra S, Kearney L;

PI Policky JL;

XX WPI; 2002-062248/08.

DR N-PSDB; AAS99894.

XX New cytoskeleton-associated proteins and polynucleotides, useful for

PT diagnosing, preventing and treating cell proliferative, autoimmune,

PT inflammatory, neurological, cell motility, reproductive and muscle

PT disorders.

XX Claim 1; Page 130-131; 194pp; English.

XX The invention relates to human cytoskeleton-associated polypeptides

CC (CYSKP) and their associated polynucleotide sequences. The sequences are

CC useful in the treatment of disorders associated with overexpression or

CC underexpression of CYSKP in a patient. The disorders include cell

CC proliferative disorders (such as cancer, actinic keratosis,

CC arteriosclerosis, cirrhosis, hepatitis and psoriasis),

CC autoimmune/inflammatory disorders (such as, asthma, atherosclerosis,

CC osteoporosis, Crohn's disease, rheumatoid arthritis, diabetes mellitus

CC and anaemia), vesicle trafficking disorders (such as

CC hypercholesterolaemia, diabetes insipidus, Grave's disease and goitre),

CC gastrointestinal disorders, prion diseases, neurological disorders (such

CC as epilepsy, stroke, cerebral neoplasms, Alzheimer's disease,

CC Huntington's disease, Parkinson's disease, amyotrophic lateral sclerosis

CC and other motor neuron disorders), cell motility disorders, reproductive

CC disorders (such as endometriosis and polycystic ovary syndrome), muscle

CC disorders (such as myocarditis, migraine, hypertension, hypoglycaemia,

CC myocardial infarction, epilepsy and muscular dystrophy), spinal cord

CC diseases, central nervous system disorders (such as Down syndrome and

CC cerebral palsy) and mental disorders (such as anxiety and schizophrenia).

CC Sequences AAU74330-AAU74363 represent human CYSKP of the invention

XX

SQ Sequence 488 AA;

Query Match 58.2%; Score 908; DB 5; Length 488;

Best Local Similarity 56.7%; Pred. No. 4.2e-89;

Matches 165; Conservative 48; Mismatches 78; Indels 0; Gaps 0;

QY 2 DITSDADAVDLTEAEWEDLTQOYYSLVHGDGFINSRNYFSQCQALLNRITSVNPOTDI 61

DB 2 DIDKDLAPLYLTPEGWSLFQRYQVHVEGAEHLRLDTQVRCEDILQQLQAVVQIDM 61

QY 62 DGLRNIWIKPAKSRGRDIVCMRDVEEILEAAADHPLSRDNKVVQKYIETPLLICDT 121

DB 62 EGDRTNWIWFKGAKSRGIMCMCHLEMLKLVNGNPVVMKDGKVVQKYIERPLLIFFT 121

QY 122 KFDIRQWFLVDNWNPLTIWFKESYLRFSQRLSKLDSAIHLGNNAVKYLNKDVGRS 181

DB 122 KFDLRQWFLVDNWNPLTIWFKESYLRFSQRLSKLDSAIHLGNNSVHLNNSIOKHLNSCHRH 181

QY 182 PLPLAHNMWTSRFOEYIQGRGAVMGSVIYFSMKKAIHAAMKVAQDHVEPRKNSPELY 241

DB 182 PLPLPPDNWSSQRFQALHQEWGAPNAWSTIIVCGMKDAVIHALQTSQDTVQCRKASPELY 241

QY 242 GADPVLGRDPRPMLIENSSPTWHPSTPTVTAQLCAQVQEDTIKAVDRSCD 292

DB 242 GADPFGEDQPMLEINASPTWAPSTAVTARLCAGVQADTLRVYIDRMLD 292

RESULT 11

AAB94796

ID AAB94796 standard; protein; 744 AA.

XX AAB94796;

AC

XX

XX 26-JUN-2001 (first entry)

XX

DE Human protein sequence SEQ ID NO:15921.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-00116126.

XX 29-JUL-1999; 99JP-00248036.

XX 27-AUG-1999; 99JP-00300253.

XX 11-JAN-2000; 2000JP-00118776.

XX 02-MAY-2000; 2000JP-00183767.

XX 09-JUN-2000; 2000JP-00241899.

XX (HELI-) HELIX RES INST.

PA Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-

PT length cDNAs defined in the specification, and for the detection and/or

PT diagnosis of the abnormality of the proteins encoded by the full-length

PT cDNAs.

XX Claim 8; SEQ ID NO 15921; 2537pp + Sequence Listing; English.

XX The present invention describes primer sets for synthesising 5602 full-

CC length cDNAs defined in the specification. Where a primer set comprises:

CC (a) an oligo-dr primer and an oligonucleotide complementary to the

CC complementary strand of a polynucleotide which comprises one of the 5602

CC nucleotide sequences defined in the specification, where the

CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination

CC of an oligonucleotide comprising a sequence complementary to the

CC complementary strand of a polynucleotide which comprises a 5'-end

CC sequence and an oligonucleotide comprising a sequence complementary to a

CC polynucleotide which comprises a 3'-end sequence, where the

CC oligonucleotide comprises at least 15 nucleotides and the combination of

CC the 5'-end sequence/3'-end sequence is selected from those defined in the

CC specification. The primer sets can be used in antisense therapy and in

CC gene therapy. The primers are useful for synthesising polynucleotides,

CC particularly full-length cDNAs. The primers are also useful for the

CC detection and/or diagnosis of the abnormality of the proteins encoded by

CC the full-length cDNAs. The primers allow obtaining of the full-length

CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and

CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893

CC represent human amino acid sequences; and AAH13629 to AAH13632 represent

CC oligonucleotides, all of which are used in the exemplification of the

CC present invention

XX

SQ Sequence 744 AA;

Query Match 58.2%; Score 908; DB 4; Length 744;

Best Local Similarity 56.7%; Pred. No. 7.8e-89;

Matches 165; Conservative 48; Mismatches 78; Indels 0; Gaps 0;

QY 2 DITSDADAVDLTEAEWEDLTQOYYSLVHGDGFINSRNYFSQCQALLNRITSVNPOTDI 61

DB 214 DIDKDLAPLYLTPEGWSLFQRYQVHVEGAEHLRLDTQVRCEDILQQLQAVVQIDM 273

QY 62 DGLRNIWIKPAKSRGRDIVCMRDVEEILEAAADHPLSRDNKVVQKYIETPLLICDT 121

DB 274 EGDRTNWIWFKGAKSRGIMCMCHLEMLKLVNGNPVVMKDGKVVQKYIERPLLIFFT 333

QY 122 KFDIRQWFLVDNWNPLTIWFKESYLRFSQRLSKLDSAIHLGNNAVKYLNKDVGRS 181

DB 334 KFDLRQWFLVDNWNPLTIWFKESYLRFSQRLSKLDSAIHLGNNSVHLNNSIOKHLNSCHRH 393

QY 182 PLLPAHNMWTSRFOEYLQQRGAVWGSVIYPSMKKATAHAKVQAQDHVEPRKNSFELY 241
 DB 394 PLLPPDNMWSQRFOAHLQEMGAPNAWSTIIIVPGMKDAVIHALQTSQDTVQCRKASFELY 453
 QY 242 GADFLVGRDRFRPWLIEINSSPTMHPSTPVTQAQCAQVQEDTIKVAVDKSCD 292
 DB 454 GADFLVGRDRFRPWLIEINSSPTMHPSTPVTQAQCAQVQEDTIKVAVDKSCD 292

RESULT 12
 ADM05524
 ID ADM05524 standard; protein; 326 AA.
 XX
 AC ADM05524;
 XX
 DT 20-MAY-2004 (first entry)
 DE Human protein of the invention SEQ ID NO:4209.
 XX
 KW human; gene therapy; diagnostic marker; pharmaceutical.
 OS Homo sapiens.
 XX
 PN EP1347046-A1.
 XX
 PD 24-SEP-2003.
 XX
 PF 12-APR-2002; 2002EP-00008400.
 XX
 PR 22-MAR-2002; 2002JP-00137785.
 XX
 PA (REAS-) RES ASSOC BIOTECHNOLOGY.
 XX
 PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
 PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
 PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
 XX
 DR WPI; 2003-723558/69.
 DR N-PSDB; ADM03081.
 XX
 PT New polynucleotides and polypeptides are useful in gene therapy, for
 PT developing a diagnostic marker or medicines for regulating their
 PT expression and activity, or as a target of gene therapy.
 XX
 PS Claim 1; SEQ ID NO 4209; 305pp; English.
 XX
 CC The invention relates to a novel human polynucleotide and the encoded
 CC polypeptide. A polynucleotide of the invention may have a use in gene
 CC therapy. An oligonucleotide of the invention ADM06202-ADM06773 is useful
 CC as a primer for synthesizing the polynucleotide or as a probe for
 CC detecting the polynucleotide. The polynucleotides ADM01316-ADM03758 are
 CC useful in gene therapy, for developing a diagnostic marker or medicines
 CC for regulating their expression and activity, or as a target of gene
 CC therapy. The proteins ADM03759-ADM06201 encoded by the polynucleotides
 CC are useful as pharmaceutical agents. The present sequence represents a
 CC protein sequence of the invention.
 XX
 SQ Sequence 326 AA;
 Query Match 58.0%; Score 905; DB 7; Length 326;
 Best Local Similarity 56.7%; Pred. No. 4.9e-89;
 Matches 165; Conservative 48; Mismatches 78; Indels 0; Gaps 0;

QY 2 DIDSADAVEDLTAEWEDLTQOYSLVHGDAFISNRYFSQCALLNRITSVNPQTDI 61
 DB 2 DIDKDLAPLYLTPEGWSLFLORYQVHGEALHDLTVQVRCEDIQLQQAQVVPQIDM 61
 QY 62 DGLRNWIKPAKSGRSDIVCMRVEEILEAAADHPLSRDNKRWVQKYIETPLLICDT 121
 DB 62 EGDRIWIVKPGAKSGRGIMCMDHLEMLKLVNGNVPVMKDGKVVQKYIERPLLIFFT 121
 QY 122 KFDIRQWFLVTDWNPPLTWTFYKESYLRSTQRFSLKLNLSVHLCCNSIQKHLNSCHRH 181

DB 122 KFDLRQWFLVTDWNPPLTWTFYKESYLRSTQRFSLKLNLSVHLCCNSIQKHLNSCHRH 181
 QY 182 PLLPAHNMWTSRFOEYLQQRGAVWGSVIYPSMKKATAHAKVQAQDHVEPRKNSFELY 241
 DB 182 PLLPPDNMWSQRFOAHLQEMGAPNAWSTIIIVPGMKDAVIHALQTSQDTVQCRKASFELY 241
 QY 242 GADFLVGRDRFRPWLIEINSSPTMHPSTPVTQAQCAQVQEDTIKVAVDKSCD 292
 DB 242 GADFLVGRDRFRPWLIEINSSPTMHPSTPVTQAQCAQVQEDTIKVAVDKSCD 292

RESULT 13
 AAB58909
 ID AAB58909 standard; protein; 362 AA.
 XX
 AC AAB58909;
 XX
 DT 27-MAR-2001 (first entry)
 DE Breast and ovarian cancer associated antigen protein sequence SEQ ID 617.
 XX
 KW Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;
 KW neotropic; neuroprotective; antiviral; antiallergic; hepatotropic;
 KW antidiabetic; antiinflammatory; antiulcer; vulnery; anticonvulsant;
 KW antibacterial; antifungal; antiparasitic; cardiant; immune disorder;
 KW Addison's disease; allergy; autoimmune haemolytic anaemia;
 KW autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
 KW multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
 KW cardiovascular disorder; wound healing; neurological disease.
 XX
 OS Homo sapiens.
 XX
 PN WO200055173-A1.
 XX
 PD 21-SEP-2000.
 XX
 PF 08-MAR-2000; 2000WO-US005881.
 XX
 PR 12-MAR-1999; 99US-0124270P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Ruben SM;
 XX
 DR WPI; 2000-611515/58.
 DR N-PSDB; AAF21812.
 XX
 PT New human breast and ovarian cancer associated gene sequences and the
 PT polypeptides encoded by these genes, useful in the prevention, treatment
 PT and diagnosis of cancer, immune disorders, cardiovascular disorders and
 PT neurological diseases.
 XX
 PS Claim 11; Page 1056-1057; 1299pp; English.
 XX
 CC Sequences AAF21614 - AAF22031 represent DNA sequences encoding human
 CC proteins AAB58711 - AAB59128. The DNA and protein sequences are
 CC associated with breast and ovarian cancer. Included in the invention are
 CC sequences AAF22032 - AAF22040 and AAB59129 which are used in the
 CC isolation and characterisation of the DNA and protein sequences of the
 CC invention. The breast and ovarian cancer associated DNA, protein, agonist
 CC or antagonist sequences exhibit cytostatic; immunosuppressive; neotropic;
 CC neuroprotective; antiviral; antiallergic; hepatotropic; antidiabetic;
 CC antiinflammatory; antiulcer; vulnery; anticonvulsant; antibacterial;
 CC antifungal; antiparasitic and cardiant activity. The polynucleotide and
 CC protein sequences are used in the diagnosis of cancer, particularly
 CC breast and ovarian cancer. The nucleic acid sequences, proteins, agonists
 CC and agonists may also be used in the diagnosis, prevention and treatment
 CC of immune disorders e.g. Addison's disease, allergies, autoimmune
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis;
 CC cardiovascular disorders such as myocardial ischaemia; wound healing;
 CC neurological diseases such as cerebral anoxia and epilepsy; and
 CC infectious diseases

XX SQ Sequence 362 AA;
Query Match 57.5%; Score 897; DB 3; Length 362;
Best Local Similarity 57.5%; Pred. No. 4.2e-88;
Matches 161; Conservative 47; Mismatches 72; Indels 0; Gaps 0;
QY 13 LTAEWEDLTQYYSLVHGDAFTNSGRNYSQOALLNRITSVNPQTDIDGLNIWIKP 72
DB 30 LTPEGHSLQRYQYVQVHEGAELRHLDTVQRCEDILQQLQAVVQIDMEGDRNIWIKP 89
QY 73 AAKSRGRDIVCMRVEIELEAAADHPLSRDNKVVQKYEITETPLLLCDTKFDIROWFLVT 132
DB 90 GANSRGRGIMCMHLEMLKLVNGVPMVKDGWVQKYEITETPLLLFGTKFDIROWFLVT 149
QY 133 DMNPLTIWFKESYLRFSTQRFSLDKLDSAIHLNNAVQKYLKNDVGRSPLLPANHWTS 192
DB 150 DMNPLTIWFKESYLRFSTQRFSLDKLDSAIHLNNAVQKYLKNDVGRSPLLPANHWTS 209
QY 193 TRFOEYLQROGRGAVGWSVITYPSMKKAIAHAMKVAQDHVEPRKNSFELYGADFLGRDFF 252
DB 210 QRFQALHQAEMGAPNAWSTIIVPGMKDAVIHALQTSQDTVQCRKASFELYGADVFGEQDF 269
QY 253 PWLIEINSSPTMHPSTPTVQALCAQVQEDTIKVAVDVRS 292
DB 270 PWLIEINSSPTMHPSTPTVQALCAQVQEDTIKVAVDVRS 309

RESULT 14
ADJ93455
ID ADJ93455 standard; protein; 292 AA.

XX AC ADJ93455;
XX DT 06-MAY-2004 (first entry)
XX DE Human HOTTL protein sequence SeqID2.

XX testis-specific tubulin tyrosine-ligase-like polypeptide;
KW BGS-42 polypeptide; cytosolic; respiratory-gen; gastrointestinal-gen;
KW neuroprotective; endocrine-gen; anti-inflammatory; anabolic; hypertensive;
KW osteoproteic; nontropic; antiparkinsonian; antiarthritic; antiasthmatic;
KW anti-HIV; antibacterial; immunosuppressive; antiseborrheic;
KW dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;
KW tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;
KW testicular disorder; testicular cancer; pulmonary disorder; lung cancer;
KW gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;
KW brain cancer; liver cancer; proliferative condition; testis; lung;
KW small intestine; brain; lymph tissue; infertility; Cushing's syndrome;
KW emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;
KW Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;
KW sepsis; acne; Sjogren's disease; scleroderma; human; HOTTL.

XX Homo sapiens.
XX WO2004005487-A2.
XX 15-JAN-2004.
XX 09-JUL-2003; 2003WO-US021605.
XX 09-JUL-2002; 2002US-0394725P.
XX (BRIM) BRISTOL-MYERS SQUIBB CO.
XX Feder JN, Wu S, Nelson TC;
XX WPI; 2004-099381/10.
XX New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,
XX useful for preventing, treating or ameliorating a medical condition, e.g.
XX aberrant cellular proliferation, reproductive disorders or testicular
XX disorders.

XX PS Disclosure; SEQ ID NO 5; 343pp; English.
XX CC This invention relates to a novel testis-specific tubulin tyrosine-ligase
CC -like polypeptide, designated the BGS-42 polypeptide. The invention may
CC be useful for the development of compounds with a cytosolic, respiratory
CC -Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,
CC antiinflammatory, anabolic, hypertensive, osteoproteic, nontropic,
CC antiparkinsonian, antiarthritic, antiasthmatic, anti-HIV, antibacterial,
CC immunosuppressive, antiseborrheic or dermatological activity acting as
CC tyrosine ligase modulators. In addition, the disclosed sequences may be
CC used for gene therapy. The BGS-42 polypeptide or polynucleotide can be
CC used for diagnosing a pathological condition or a susceptibility to a
CC pathological condition in a subject, and for preventing, treating or
CC ameliorating a medical condition, such as a disorder related to aberrant
CC tubulin ligase activity, a disorder related to aberrant tubulin-
CC carboxypeptidase activity, aberrant cellular proliferation, reproductive
CC disorders, testicular disorders, testicular cancer, pulmonary disorders,
CC lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
CC neural disorders, brain cancer, liver cancer, or proliferative condition
CC of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
CC polypeptide, polynucleotide, or their modulators are also useful for
CC treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
CC disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
CC -42 polypeptide can be used as a preventive agent for immunological
CC disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
CC disease or scleroderma. The antibodies may be used to purify, detect and
CC target the BGS-42 polypeptides. The present sequence is that of the human
CC HOTTL protein which is related to the invention. Note: The present
CC sequence does not appear in the specification but was obtained from
CC Genbank.
XX SQ Sequence 292 AA;

Query Match 52.4%; Score 817; DB 8; Length 292;
Best Local Similarity 62.5%; Pred. No. 1.5e-79;
Matches 145; Conservative 37; Mismatches 50; Indels 0; Gaps 0;

QY 61 IDGRNIWIKPAKSRGRDIVCMRVEIELEAAADHPLSRDNKVVQKYEITETPLLICD 120
DB 1 MEGDRFWIIVKPGAKSRGIMCMHLEMLKLVNGVPMVKDGWVQKYEITETPLIFG 60
QY 121 TKEDIRQWFLVTDMNPLTIWFKESYLRFSTQRFSLDKLDSAIHLNNAVQKYLKNDVGR 180
DB 61 TKEDIRQWFLVTDMNPLTIWFKESYLRFSTQRFSLDKLDSAIHLNNAVQKYLKNDVGR 120
QY 181 SPILLPAHNMWTSRFOEYLQROGRGAVGWSVITYPSMKKAIAHAMKVAQDHVEPRKNSFEL 240
DB 121 HPLLPDNNMSSQRFQALHQAEMGAPNAWSTIIVPGMKDAVIHALQTSQDTVQCRKASFEL 180
QY 241 YGADFVLGRDFFRWLIEINSSPTMHPSTPTVQALCAQVQEDTIKVAVDVRS 292
DB 181 YGADFVFGEDFQWFLIEINASSPTMAPSTAVTARLCAGVQADTLRVRIDRLMD 232

RESULT 15
ABB65645
ID ABB65645 standard; protein; 992 AA.

XX ABB65645;
XX AC ABB65645;
XX DT 26-MAR-2002 (first entry)
XX Drosophila melanogaster polypeptide SEQ ID NO 23727.
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX Drosophila melanogaster.
XX WO200171042-A2.
XX 27-SEP-2001.

THIS PAGE BLANK (USPTO)


```
Db      125 EREAGLEAAKDFPKTPEMCEYHFLVEERKNGEITWIMKPVARSOGKGLFPRRLK 184
QY      161 EILELAADHPLSRDNK-----WVQKYIETPLLCDTKFDIRQWFLVTDWNPWTIW 213
Db      185 DIVDW-RKTRSDDDQKDDIPVENYVAQRIENPYLIGGRKFDLRYVVLVMS-----V 236
QY      214 YKESYLRFTQRFSLDKLSAHLCCNNAVOKYLKNDVGRSPLLPAIN-----MWTSTRFQ 268
Db      237 FAECLLWSGHR-----QDVHLTNVAVQK-----TSPDYHPKKGCKWTLQRF 279
QY      269 EYL-QROGRCAVNGSVIYPSMKAIAHAMKVAQDHVEPRKNSPELYGADPVLGRDPRML 327
Db      280 QYLASHKGPEAV--ETLFRDIDNIFVKSLSQSVQKVIISDKHCPELYGYDILIDQDLKPWL 337
QY      328 IEINSSPTWHPSPPTVTAQLCAQVOEDTIKVA-----VDRSCDIGNFELLWRQPVVE--- 378
Db      338 LEVNASPLTASSQEDYELKTCLEDTLHVDMEARLTGREKRVGGFDLMWMDGPFVSREE 397
QY      379 -PPFSGSGLCVAGSV-----RRARRQVLPVCNL-----KASA 411
Db      398 GAPDLGSMGNFVNTNTHLGCVNDKQKQLRQLF--CSLQVQKQASS 439
```

RESULT 2

```
US-11-072-512-3371
; Sequence 3371, Application US/11072512
; Publication No. US20060029945A1
```

GENERAL INFORMATION:

```
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: TRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOKYU
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: Novel full length cdna
; FILE REFERENCE: 084335-0191
; CURRENT APPLICATION NUMBER: US/11/072,512
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: US 60/350,978
; PRIOR FILING DATE: 2002-01-25
; PRIOR APPLICATION NUMBER: JP 2001-379298
; PRIOR FILING DATE: 2001-11-05
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3371
; LENGTH: 592
; TYPE: PRT
; ORGANISM: Homo sapiens
```

```
US-11-072-512-3371
```

```
Query Match      9.5%; Score 271.5; DB 7; Length 592;
Best Local Similarity 29.8%; Pred. No. 4.1e-15;
Matches 72; Conservative 43; Mismatches 80; Indels 47; Gaps 8;
```

```
QY      113 YFSQCQALLNRITSNVNPQTDIDGLRNITWITKPAAKSRGRDIVCMRDRVEEILELAADHPL 172
Db      195 YFQERQMLGTK-----HSYWI CKPAELSRGRGILIFSDFKDFI----- 232
QY      173 SRDNKVVQKYIETPLLCDTKFDIRQWFLVTDWNPWTIWFKESYLRFSRQFSLDKL- 231
```

```
Db      233 -FDDMYIVQKYINPLLIIGRYKCDLRIYVCVTGFKPLTIYVYQEGLVREPAKEFDLSNLQ 291
QY      232 DSAIHLCCNNAVQK-----YLNKDVGSRSPLLPAHNMWTSRFOEYLQROGRGAVM---GS 282
Db      292 NNVAHLTNSSINKSGASGYEKIEVIGHG-----CKWTLRSRFFSYLRS-----WDVDDL 339
QY      283 VIYPSMKAIAHAMKVAQDHVEPRKNSPELYGADPVLGRDPRFWLLEINSSP--TMHPST 340
Db      340 LLMWKIHRWVILITILAIAPSVFAANCPELFGFDLIDNLLKFWLLEVNYSPLATLDCST 399
QY      341 PV 342
Db      400 DV 401
```

RESULT 3

```
US-11-072-512-3621
; Sequence 3621, Application US/11072512
; Publication No. US20060029945A1
```

GENERAL INFORMATION:

```
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: TRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOKYU
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: Novel full length cdna
; FILE REFERENCE: 084335-0191
; CURRENT APPLICATION NUMBER: US/11/072,512
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: US 60/350,978
; PRIOR FILING DATE: 2002-01-25
; PRIOR APPLICATION NUMBER: JP 2001-379298
; PRIOR FILING DATE: 2001-11-05
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3621
; LENGTH: 404
; TYPE: PRT
; ORGANISM: Homo sapiens
```

```
US-11-072-512-3621
```

```
Query Match      6.4%; Score 183; DB 7; Length 404;
Best Local Similarity 25.5%; Pred. No. 8.6e-08;
Matches 82; Conservative 38; Mismatches 118; Indels 84; Gaps 17;
```

```
QY      6 LKVVVSHQSCRSRSRKPDRQREAGSSDLSSRQDAENAEAK-----LRGLPGQLVDIA 59
Db      119 LKW-----CEVKSRSDSYGSFRE--GBQLLYQLPNNKLLTTKIGLLSTLRG--RARAMSRA 169
QY      60 CKVQAYLQGLQEHEDITSADAVEDLTEAWEDLTQYYISLVHGDAFISNRNYFSQCOA 119
Db      170 SKVPGGVQARLEK-----DAAAPALEDL--PW-----TSPGYLRPQRV 205
QY      120 LLNRITSVNPQT---DIDGLR-----NIWIKPAAKSRGRDIVCMRDRVEEILELA 166
Db      206 L--RMBEFFPFTVRLDLKHREAFFTLFDETOIWKCPASNOQKGIFLLRNQBEVAALQ 263
QY      167 A-----ADHPLSRDNKW-----VVQKYIETPLLCDTKFDIRQWFLVTDWNPWTIWFK 215
Db      264 AKTRSMEDDPIHHKTPFRGQARVVQRYIQNPLLVDCGRKFDVRSYLLIACCTPYMI--PFG 322
```


Db 298 AAPCPPELPAPALLPAPCPRA-----LLPLCAYGAGEPA---RLGAREAEVPTTAPPLLLA 350

QY 458 PLRGAAGSQAQPTRTKAAGKVELPACPCRHVDSQAPNTGVFVA-OPAKSMDPNQINAH 516

Db 351 PLPAAA-----PAKPLRGAAGGAHL-YCPLR-----LPAALQAALVRRPGPHLSY 395

QY 517 PLEPVL 522

Db 396 PVETLL 401

RESULT 7

US-11-072-512-3402

; Sequence 3402, Application US/11072512

; Publication No. US20060029945A1

; GENERAL INFORMATION:

; APPLICANT: ISOGAI, TAKAO

; APPLICANT: SUGIYAMA, TOMOYASU

; APPLICANT: OTSUKI, TETSUJI

; APPLICANT: WAKAMATSU, AI

; APPLICANT: SATO, HIROYUKI

; APPLICANT: ISHII, SHIZUKO

; APPLICANT: YAMAMOTO, JUN-ICHI

; APPLICANT: ISONO, YUUKO

; APPLICANT: HIO, YURI

; APPLICANT: OTSUKA, KAORU

; APPLICANT: NAGAI, KEIICHI

; APPLICANT: IRIE, RYOTARO

; APPLICANT: TAMECHIKA, ICHIRO

; APPLICANT: SEKI, NAOHICO

; APPLICANT: YOSHIKAWA, TSUTOMU

; APPLICANT: OTSUKA, MOTOYUKI

; APPLICANT: NAGAHARI, KENJI

; APPLICANT: MASUHO, YASUHIKO

; TITLE OF INVENTION: Novel full length cdna

; FILE REFERENCE: 084335-0191

; CURRENT APPLICATION NUMBER: US/11/072,512

; PRIOR FILING DATE: 2005-03-07

; PRIOR APPLICATION NUMBER: US 60/350,978

; PRIOR FILING DATE: 2002-01-25

; PRIOR APPLICATION NUMBER: JP 2001-379298

; PRIOR FILING DATE: 2001-11-05

; NUMBER OF SEQ ID NOS: 4096

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 3402

; LENGTH: 710

; TYPE: PRT

; ORGANISM: Homo sapiens

US-11-072-512-3402

Query Match 3.6%; Score 103; DB 7; Length 710;

Best Local Similarity 23.7%; Pred. No. 1.3;

Matches 58; Conservative 20; Mismatches 79; Indels 88; Gaps 11;

QY 325 PWLEINS-----SPTMHPSTPTAQLCAQVEDTIKVAVDSCDIGNFELLWRQVVEPPP 381

Db 377 PTLIALNSALLAATLDPPSGTTPQPC-----VLSAPQPGPT 413

QY 382 FS-----GSLCVAGSVVRARROVLPVCNLKASASILDAAQLKARGPSAMPDPAQ 432

Db 414 SSVTTATTDPGASLGLKAPNSGRPPQLLSPL-----LGASLLGLSLTSPPGALPSLLQ 469

QY 433 GPPSPALQDRLGKEEGLPLALLAPLRGAAGSQAQPTRTKAAGV-----480

Db 470 -PPGILLSGQLGLO-----LLP-----GGAPPLPSEASSPLACLQLSLQIIP 512

QY 481 ELPACPCRHVDSQAPNTGVVQAQPAKSWDPNQLNAHLEPVLRLGLKTAEGALRPPP-----536

Db 513 EQPEAPC-----LPPESPASALEP-----EPARPLSALAPPHGSPDPVPPELL 556

QY 537 GKGGS 541

||:|:

Db 557 TGRGS 561

RESULT 8

US-11-072-512-2246

; Sequence 2246, Application US/11072512

; Publication No. US20060029945A1

; GENERAL INFORMATION:

; APPLICANT: ISOGAI, TAKAO

; APPLICANT: SUGIYAMA, TOMOYASU

; APPLICANT: OTSUKI, TETSUJI

; APPLICANT: WAKAMATSU, AI

; APPLICANT: SATO, HIROYUKI

; APPLICANT: ISHII, SHIZUKO

; APPLICANT: YAMAMOTO, JUN-ICHI

; APPLICANT: ISONO, YUUKO

; APPLICANT: OTSUKA, KAORU

; APPLICANT: NAGAI, KEIICHI

; APPLICANT: IRIE, RYOTARO

; APPLICANT: TAMECHIKA, ICHIRO

; APPLICANT: SEKI, NAOHICO

; APPLICANT: YOSHIKAWA, TSUTOMU

; APPLICANT: OTSUKA, MOTOYUKI

; APPLICANT: NAGAHARI, KENJI

; APPLICANT: MASUHO, YASUHIKO

; TITLE OF INVENTION: Novel full length cdna

; FILE REFERENCE: 084335-0191

; CURRENT APPLICATION NUMBER: US/11/072,512

; CURRENT FILING DATE: 2005-03-07

; PRIOR APPLICATION NUMBER: US 60/350,978

; PRIOR FILING DATE: 2002-01-25

; PRIOR APPLICATION NUMBER: JP 2001-379298

; PRIOR FILING DATE: 2001-11-05

; NUMBER OF SEQ ID NOS: 4096

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 2246

; LENGTH: 723

; TYPE: PRT

; ORGANISM: Homo sapiens

US-11-072-512-2246

Query Match 3.6%; Score 101.5; DB 7; Length 723;

Best Local Similarity 23.0%; Pred. No. 1.8;

Matches 65; Conservative 38; Mismatches 115; Indels 65; Gaps 14;

QY 296 MKVAQDHVEPRKNSFE-----LYGADFVLGRDRFPW-----LIEINSSPTMH 337

Db 58 MKATQSYLQEKVNSFQNMKNQYMGNSMPPFPVYSQMPPLPTMPPVLPPLPFPVMP 117

QY 338 PSTPVTQAQ---LCAQVQEDTIKVAVDSCDIGNFELLWRQVVEPRPFGSDLCVAGVSV 394

Db 118 PALPATVPPPGMPPPPVPMPPSLPTSVPPPGMPPSLSSAGAPPPVLPFP-----SLFSAGPP- 171

QY 395 RRARROVLPVCNLKASA--SLLDAQPL-KARGPSAMPDPA--QG-----PPSPALQR 441

Db 172 -----PVLPPPSLSSSTAPPVPMPLPPLSSSATPPGIPPGVPGQIIPQLTAAPVPPSSS 226

QY 442 DLGLKEEKGFLALLAPLRGAAGSQAQPTR-----TKAAGKVELPACPCRHVDSQAP-- 495

Db 227 QSSQVPEKPRALLPTEV-----SFGSAPPTTYHPPLQSAG-----PSEQVNSKAPLS 274

QY 496 NTGVVQAQPAKSWDPNQLNAHLEPV--LRGLKTAEGALRPPP 536

Db 275 KSAIPYSFSSDOGLGESSAAPQPIITAVKDMPVRSGLLPDP 317

RESULT 9

US-11-072-512-3151

; Sequence 3151, Application US/11072512

; Publication No. US20060029945A1

; GENERAL INFORMATION:

; APPLICANT: ISOGAI, TAKAO


```
/ GENERAL INFORMATION:
/ APPLICANT: Tang, Y Tom
/ APPLICANT: Liu, Chenghua
/ APPLICANT: Asundi, Vinod
/ APPLICANT: Chen, Rui-hong
/ APPLICANT: Qian, Xiaohong B.
/ APPLICANT: Wang, Zhiwei
/ APPLICANT: Wehrman, Tom
/ APPLICANT: Zhang, Jie
/ APPLICANT: Zhou, Ping
/ APPLICANT: Cao, Yi-Cheng
/ APPLICANT: Dmanac, Radoje T.
/ TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
/ FILE REFERENCE: 785CIP4CN
/ CURRENT APPLICATION NUMBER: US/11/000,463
/ CURRENT FILING DATE: 2004-11-29
/ PRIOR APPLICATION NUMBER: 10/291,265
/ PRIOR FILING DATE: 2002-11-08
/ PRIOR APPLICATION NUMBER: PCT/US01/02623
/ PRIOR FILING DATE: 2001-01-25
/ PRIOR APPLICATION NUMBER: 09/922,279
/ PRIOR FILING DATE: 2001-08-03
/ PRIOR APPLICATION NUMBER: 09/491,404
/ PRIOR FILING DATE: 2000-01-25
/ PRIOR APPLICATION NUMBER: 09/617,746
/ PRIOR FILING DATE: 2000-07-17
/ PRIOR APPLICATION NUMBER: 09/631,451
/ PRIOR FILING DATE: 2000-08-03
/ PRIOR APPLICATION NUMBER: 09/633,870
/ PRIOR FILING DATE: 2000-09-15
/ NUMBER OF SEQ ID NOS: 944
/ SOFTWARE: Fast-SEQ for Windows Version 3.0
/ SEQ ID NO 375
/ LENGTH: 423
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-11-000-463-375
```

```
Query Match 3.5%; Score 99.5; DB 7; Length 423;
Best Local Similarity 24.4%; Pred. No. 1.3;
Matches 89; Conservative 34; Mismatches 125; Indels 117; Gaps 22;

Qy 181 QKIETPLLCDFKFDTRQWFLVTD-----WNPLTWFY----KSYLRFSTQRFSLDKLDS 233
Db 128 QKLFEKILKRDVAHKV---FATTSIKSFQQLNLYGFRKRRQCTFRTFTRIFSAGELVS 184

Qy 234 AIHLNNNAVQKLVNDVGRSPFLPAH-----NMWTSRFOYLQGRGVWGSQVIYPSM 288
Db 185 ILNKLEFYCHPYFORD---SPHLLVRMKRRVGVKSAPRHQEDKPEAAGS-----231

Qy 289 KKAIAHAMKVAQDHVEPRKNSFELYGADFVLGRDRFP-----WLIEINSSPTMHPSTPVT 344
Db 232 --CLAPADTQODHTSPNEN-----DQVTPQHREPAQNPNTQIRSGSAP---PATPVWV 279

Qy 345 QLCQAOVEDTIKAVDRSCDIGNPELLWROQVVEPPPPSGSDLCVAGSVRRARRQVLVP 404
Db 280 PDSA-----VASDNS-----PVTQP-----AGEWSEGSQAHTVPV 309

Qy 405 CNLKASASLLDAQPKARGSPAMP-----DPAQ-----GP--PSPALQDRLGLKEEGLP 452
Db 310 A-----AVP-----GPNALPFLVYVPGSPTQNSYGPVVALPTASRSTLAMDTTGLP 355

Qy 453 LALLAP-----LRGAASGGAAQPTRTKAAGKV---ELPA-----CPCRHVDSQ--APNTGV 499
Db 356 APGMLPFCHLWVPVTLVAAGAAQP-----AASMVNFPHLPALHHCHPHSHRTSQYWPASDG 411

Qy 500 PVAQP 504
Db 412 PQAYP 416
```

RESULT 12
US-11-075-185-6

```
/ Sequence 6, Application US/11075185
/ Publication No. US20050266434A1
/ GENERAL INFORMATION:
/ APPLICANT: REEVES, CHRISTOPHER D
/ APPLICANT: JULIEN, BRYAN
/ APPLICANT: REID, RALPH
/ TITLE OF INVENTION: BIOSYNTHETIC GENE CLUSTER FOR AMBRUTICINS
/ FILE REFERENCE: 010099.03
/ CURRENT APPLICATION NUMBER: US/11/075,185
/ CURRENT FILING DATE: 2005-03-07
/ PRIOR APPLICATION NUMBER: US 60/551,103
/ PRIOR FILING DATE: 2004-03-08
/ PRIOR APPLICATION NUMBER: US 60/568,290
/ PRIOR FILING DATE: 2004-05-04
/ NUMBER OF SEQ ID NOS: 61
/ SOFTWARE: PatentIn version 3.3
/ SEQ ID NO 6
/ LENGTH: 1892
/ TYPE: PRT
/ ORGANISM: Sorangium cellulosum
US-11-075-185-6
```

```
Query Match 3.5%; Score 99.5; DB 7; Length 1892;
Best Local Similarity 20.4%; Pred. No. 11;
Matches 133; Conservative 70; Mismatches 221; Indels 229; Gaps 30;

Qy 20 RSKPRDQREBA-----GSSDLSRQDAENAEAKLRGLP-----52
Db 1154 QSLARLSREBAAGDPVLVPFAWTVALYATGATELRARIALQASG--GAPAVASLLL 1210

Qy 53 ---GQLVDIACKVCOAYLQLEHEDIDTSADAVEDLTAEWEDL-----TQYYSLV 101
Db 1211 ADAHGRVTATTVGRVGAQAQTR-----SAASRAEPMYRVAVTVDVALEATWAEHVL 1265

Qy 102 HGDAFISNSRNYPSCQALLNRITSVNPQTDIDGLRNIWIKPAKSRGRDIVCMQDRVEE 161
Db 1266 GGDGALA-----AALGVRAAAGLP-LEALAD-----GAAAPR-----RL 1300

Qy 162 ILBLAAD-----HPLSRDNKVVQKVIETPLLICDTKFDIRQWFLVTD-----W 206
Db 1301 VVDLTAGDAGAVVAHVAAVVGALALVQGWLAAPQLAATELLVVTTRCAVATGDEGVDA 1360

Qy 207 NPLTIWFYKESYLRFSSTQRFSLDKLSAHLNNNAVQKLVNDVGRSPFLPAHNMWTSR 266
Db 1361 GPAVW-----GLLRATRAEYP-----DRAVRVL-----DVGREPLDGA-----1394

Qy 267 FBYLQROGRGAVWGSVIYPSMKKAIHAHMKVAQ-----DHVEPRKNSFELYGA 315
Db 1395 ----LLRRALAA--GTEPELSVRSGEARAPRLREVRSSEPAAPATRLDPDGTAL-ITGG 1447

Qy 316 DFVLGRDRFPWLIEINSSPTMH-----PSTPVTQALCAQVQ---EDTIKVAVD 360
Db 1448 TGBLGRHVAKHLV-----TAHGVVRHLVLTSSRRGMDAPDAALVDELRAAGAAATVDVA--1499

Qy 361 RSCDIGNFELLWQPVVEPPPSG-----SDLCVAGSVVRRARRQVLV-----403
Db 1500 -ACDAADAAL--AAVVEALPAARPLTAVVHTTAGVLDDSVVTKLSAQRLARVLPKVDGA 1556

Qy 404 -----VCNLKASASLLD--AQPLKARGSPAMPDPAQ 432
Db 1557 FHLHELTKHAPLAAFVLFSSAAGTGLSGPQANVAAANTFLDALASHLRARG-----1607

Qy 433 GPSPALQDRLGLKEEGLPLALLAPLRGAESGGAQPTRTKAAGKVLP-ACPCRHVD 491
Db 1608 ---VPMSLAWGEFWAQTGL-----GMTAHLGAADIARMKRGHVSVMPVAGQLRLD 1655

Qy 492 ---SQAPNTGVPVPAQPAKSWDPNQLNAHPLPVLRLGLKTAEGALRPPPGGKGS 541
Db 1656 RALAQAETLVPLALDLSLQKAGSNAGPVPFLLRGLVRAPAGRRTAASAGA 1708
```

RESULT 13
US-11-096-568A-21828

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2006, 12:55:16 ; Search time 77.9322 Seconds
(without alignments)
2900.544 Million cell updates/sec

Title: US-10-635-977-2
Perfect score: 2854
Sequence: 1 MASSILKVVVSHQSCSRSSR.....LRGLKTAEALRPPPGKXGS 541

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_AA_Main:*
1: /cgn2_6/prodata/1/pubpaa/US07_PUBCOMB.pap:*
2: /cgn2_6/prodata/1/pubpaa/US08_PUBCOMB.pap:*
3: /cgn2_6/prodata/1/pubpaa/US09_PUBCOMB.pap:*
4: /cgn2_6/prodata/1/pubpaa/US10A_PUBCOMB.pap:*
5: /cgn2_6/prodata/1/pubpaa/US10B_PUBCOMB.pap:*
6: /cgn2_6/prodata/1/pubpaa/US11_PUBCOMB.pap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	2854	100.0	541	4	US-10-615-659-2
2	2854	100.0	541	4	US-10-635-977-2
3	2843	99.6	541	4	US-10-615-659-13
4	2843	99.6	541	4	US-10-635-977-13
5	1560	54.7	293	4	US-10-615-659-14
6	1560	54.7	293	4	US-10-635-977-14
7	1299	45.5	242	4	US-10-615-659-4
8	1299	45.5	242	4	US-10-635-977-4
9	1013	35.5	488	4	US-10-275-595A-5
10	969.5	34.0	362	3	US-09-925-298-617
11	969.5	34.0	362	4	US-10-102-806-617
12	956.5	33.5	352	4	US-10-615-659-7
13	956.5	33.5	352	4	US-10-635-977-7
14	956.5	33.5	352	5	US-10-756-149-5267
15	944.5	33.1	326	4	US-10-108-260A-4209
16	865.5	30.3	292	4	US-10-615-659-5
17	865.5	30.3	292	4	US-10-635-977-5
18	722.5	25.3	992	6	US-11-097-143-23727
19	536.5	18.8	719	6	US-11-097-143-23415
20	375.5	13.2	566	5	US-10-450-763-36330
21	328	11.5	496	6	US-11-097-143-13014
22	322	11.3	330	4	US-10-615-659-26
23	322	11.3	330	4	US-10-635-977-26
24	303.5	10.6	1226	5	US-10-756-149-5184
25	302	10.6	268	4	US-10-424-599-205823
26	299.5	10.5	423	4	US-10-615-659-8
27	299.5	10.5	423	4	US-10-635-977-8

28	297.5	10.4	827	6	US-11-097-143-9312	Sequence 9312, Ap
29	291	10.2	989	6	US-11-097-143-7680	Sequence 7680, Ap
30	282.5	9.9	487	6	US-11-097-143-4080	Sequence 4080, Ap
31	280.5	9.8	439	4	US-10-104-047-3658	Sequence 3658, Ap
32	274	9.6	49	4	US-10-615-659-21	Sequence 21, Appl
33	274	9.6	49	4	US-10-615-659-22	Sequence 22, Appl
34	274	9.6	49	4	US-10-635-977-21	Sequence 21, Appl
35	274	9.6	49	4	US-10-635-977-22	Sequence 22, Appl
36	271.5	9.5	524	3	US-09-864-761-38213	Sequence 38213, A
37	271.5	9.5	592	4	US-10-104-047-3371	Sequence 3371, Ap
38	270	9.5	917	6	US-11-097-143-21303	Sequence 21303, A
39	259	9.1	379	4	US-10-615-659-6	Sequence 6, Appl
40	259	9.1	379	4	US-10-635-977-6	Sequence 6, Appl
41	258	9.0	92	4	US-10-424-599-262294	Sequence 262294,
42	254.5	8.9	377	4	US-10-210-130-1118	Sequence 118, App
43	254.5	8.9	377	4	US-10-250-613-1	Sequence 1, Appl
44	237	8.3	42	4	US-10-615-659-24	Sequence 24, Appl
45	237	8.3	42	4	US-10-635-977-24	Sequence 24, Appl

ALIGNMENTS

RESULT 1

US-10-615-659-2
; Sequence 2, Application US/10615659
; Publication No. US20040157234A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; CURRENT FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 541
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-615-659-2

Query Match	100.0%	Score 2854;	DB 4;	Length 541;
Best Local Similarity	100.0%	Pred. No. 3.1e-229;	Mismatches 0;	Indels 0; Gaps 0;
Matches 541;	Conservative 0;			
Qy	1	MASSILKVVVSHQSCSRSSRSPKPRDQREAGSSDLSSRQDAENAEAKLRGLPGQLVDIAC	60	
Db	1	MASSILKVVVSHQSCSRSSRSPKPRDQREAGSSDLSSRQDAENAEAKLRGLPGQLVDIAC	60	
Qy	61	KVCQAYLQLEHEDITSDADAVEDLTEAEWEDLTQYYSLVHGDAFISNSRYFSCQAL	120	
Db	61	KVCQAYLQLEHEDITSDADAVEDLTEAEWEDLTQYYSLVHGDAFISNSRYFSCQAL	120	
Qy	121	LNRITSVNPQTDIDGLRNIIWKPAKSRGRDIVCMRDVEELEAAADHPLSRNKKVV	180	
Db	121	LNRITSVNPQTDIDGLRNIIWKPAKSRGRDIVCMRDVEELEAAADHPLSRNKKVV	180	
Qy	181	QKIETPLLCIDTKFDIRQWFLVTDWNPFTTWFKYESYLRFSTQRFSLDKLDSAIHLCNN	240	
Db	181	QKIETPLLCIDTKFDIRQWFLVTDWNPFTTWFKYESYLRFSTQRFSLDKLDSAIHLCNN	240	
Qy	241	AVQKYLKNDVGRSPLLPANNMTSTRFQYLRQGRGAVGWSVIYPSMKKAIAHAMKVAQ	300	
Db	241	AVQKYLKNDVGRSPLLPANNMTSTRFQYLRQGRGAVGWSVIYPSMKKAIAHAMKVAQ	300	
Qy	301	DHVEPRKNSFELYGADFVLGRDFRPWLTEINSSPTMHPSTPTAQLCAQVQEDTIKAVD	360	
Db	301	DHVEPRKNSFELYGADFVLGRDFRPWLTEINSSPTMHPSTPTAQLCAQVQEDTIKAVD	360	
Qy	361	RSCDIGNPELLWRQPVVPPPPFSGSDLCVAGSVRRARRQVLPVCNKLKASALLDAQPLK	420	

Db 361 RSCDIGNFELLWQPVVEPPFSGSDLCVAGSVRRARRQVLPVCNLKASASLLDAQPLK 420
Qy 421 ARGPSAMPDPAQGPSPALQDRLGLKEEKGFLPALLAPLRGAESGAAQPTRTKAAGKV 480
Db 421 ARGPSAMPDPAQGPSPALQDRLGLKEEKGFLPALLAPLRGAESGAAQPTRTKAAGKV 480
Qy 481 ELPACPCRHVDSQAPNTGVPVAQPAKSWDPNQLNAHPLEPVLRLGLKTAEALRPPPGKG 540
Db 481 ELPACPCRHVDSQAPNTGVPVAQPAKSWDPNQLNAHPLEPVLRLGLKTAEALRPPPGKG 540
Qy 541 S 541
Db 541 S 541

RESULT 2
US-10-635-977-2
; Sequence 2, Application US/10635977
; Publication No. US20040171131A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; TITLE OF INVENTION: TYROSINE-LIGASE-LIKE PROTEIN, BGS42
; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635,977
; PRIOR FILING DATE: 2003-08-07
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S. 10/615,659
; PRIOR FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 541
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-635-977-2

Query Match 100.0%; Score 2854; DB 4; Length 541;
Best Local Similarity 100.0%; Pred. No. 3.1e-229;
Matches 541; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MASSILKVVVSHQSCSRSSRSKPRDQREAEAGSSDLSSRQDAENAEAKLGLPGQLVDIAC 60
Db 1 MASSILKVVVSHQSCSRSSRSKPRDQREAEAGSSDLSSRQDAENAEAKLGLPGQLVDIAC 60
Qy 61 KVCQAYLGQLEHEDIDTSADAVEDLTEAEWEDLTQQYYSLVHGDVAFISNSRNYFSQCOAL 120
Db 61 KVCQAYLGQLEHEDIDTSADAVEDLTEAEWEDLTQQYYSLVHGDVAFISNSRNYFSQCOAL 120
Qy 121 LNRTSVNPTQDIDGLRNIWIIPKAAKSRGRDIVCMRVEEILELAADHPLSRDNKVV 180
Db 121 LNRTSVNPTQDIDGLRNIWIIPKAAKSRGRDIVCMRVEEILELAADHPLSRDNKVV 180
Qy 181 QKYIETPLLCIDTKFDIRQWFLVTDWNPFTIWFYKESYLPSTQRFSLDKLDSAIHLCNN 240
Db 181 QKYIETPLLCIDTKFDIRQWFLVTDWNPFTIWFYKESYLPSTQRFSLDKLDSAIHLCNN 240
Qy 241 AVQKYLKNDVGRSPLLPAHNNMTSTRFOEYLQRGVAVGWSVIYPSMKKAIHAAMKVAQ 300
Db 241 AVQKYLKNDVGRSPLLPAHNNMTSTRFOEYLQRGVAVGWSVIYPSMKKAIHAAMKVAQ 300
Qy 301 DHVEPRKNSFELYGADFVLGRDFRPWLIIEINSSPTMHPSTPTVAQLCAQVOEDTIKVAVD 360
Db 301 DHVEPRKNSFELYGADFVLGRDFRPWLIIEINSSPTMHPSTPTVAQLCAQVOEDTIKVAVD 360
Qy 361 RSCDIGNFELLWQPVVEPPFSGSDLCVAGSVRRARRQVLPVCNLKASASLLDAQPLK 420
Db 361 RSCDIGNFELLWQPVVEPPFSGSDLCVAGSVRRARRQVLPVCNLKASASLLDAQPLK 420
Qy 421 ARGPSAMPDPAQGPSPALQDRLGLKEEKGFLPALLAPLRGAESGAAQPTRTKAAGKV 480

Db 421 ARGPSAMPDPAQGPSPALQDRLGLKEEKGFLPALLAPLRGAESGAAQPTRTKAAGKV 480
Qy 481 ELPACPCRHVDSQAPNTGVPVAQPAKSWDPNQLNAHPLEPVLRLGLKTAEALRPPPGKG 540
Db 481 ELPACPCRHVDSQAPNTGVPVAQPAKSWDPNQLNAHPLEPVLRLGLKTAEALRPPPGKG 540
Qy 541 S 541
Db 541 S 541
RESULT 3
US-10-615-659-13
; Sequence 13, Application US/10615659
; Publication No. US20040157234A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; TITLE OF INVENTION: TYROSINE-LIGASE-LIKE PROTEIN, BGS42
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; PRIOR FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 13
; LENGTH: 541
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-615-659-13
Query Match 99.6%; Score 2843; DB 4; Length 541;
Best Local Similarity 99.6%; Pred. No. 2.5e-228;
Matches 539; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 MASSILKVVVSHQSCSRSSRSKPRDQREAEAGSSDLSSRQDAENAEAKLGLPGQLVDIAC 60
Db 1 MASSILKVVVSHQSCSRSSRSKPRDQREAEAGSSDLSSRQDAENAEAKLGLPGQLVDIAC 60
Qy 61 KVCQAYLGQLEHEDIDTSADAVEDLTEAEWEDLTQQYYSLVHGDVAFISNSRNYFSQCOAL 120
Db 61 KVCQAYLGQLEHEDIDTSADAVEDLTEAEWEDLTQQYYSLVHGDVAFISNSRNYFSQCOAL 120
Qy 121 LNRTSVNPTQDIDGLRNIWIIPKAAKSRGRDIVCMRVEEILELAADHPLSRDNKVV 180
Db 121 LNRTSVNPTQDIDGLRNIWIIPKAAKSRGRDIVCMRVEEILELAADHPLSRDNKVV 180
Qy 181 QKYIETPLLCIDTKFDIRQWFLVTDWNPFTIWFYKESYLPSTQRFSLDKLDSAIHLCNN 240
Db 181 QKYIETPLLCIDTKFDIRQWFLVTDWNPFTIWFYKESYLPSTQRFSLDKLDSAIHLCNN 240
Qy 241 AVQKYLKNDVGRSPLLPAHNNMTSTRFOEYLQRGVAVGWSVIYPSMKKAIHAAMKVAQ 300
Db 241 AVQKYLKNDVGRSPLLPAHNNMTSTRFOEYLQRGVAVGWSVIYPSMKKAIHAAMKVAQ 300
Qy 301 DHVEPRKNSFELYGADFVLGRDFRPWLIIEINSSPTMHPSTPTVAQLCAQVOEDTIKVAVD 360
Db 301 DHVEPRKNSFELYGADFVLGRDFRPWLIIEINSSPTMHPSTPTVAQLCAQVOEDTIKVAVD 360
Qy 361 RSCDIGNFELLWQPVVEPPFSGSDLCVAGSVRRARRQVLPVCNLKASASLLDAQPLK 420
Db 361 RSCDIGNFELLWQPVVEPPFSGSDLCVAGSVRRARRQVLPVCNLKASASLLDAQPLK 420
Qy 421 ARGPSAMPDPAQGPSPALQDRLGLKEEKGFLPALLAPLRGAESGAAQPTRTKAAGKV 480
Db 421 ARGPSAMPDPAQGPSPALQDRLGLKEEKGFLPALLAPLRGAESGAAQPTRTKAAGKV 480
Qy 481 ELPACPCRHVDSQAPNTGVPVAQPAKSWDPNQLNAHPLEPVLRLGLKTAEALRPPPGKG 540
Db 481 ELPACPCRHVDSQAPNTGVPVAQPAKSWDPNQLNAHPLEPVLRLGLKTAEALRPPPGKG 540
Qy 541 S 541

Db 541 S 541

RESULT 4

US-10-635-977-13
; Sequence 13, Application US/10635977
; Publication No. US20040171131A1

GENERAL INFORMATION:

APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635,977
; CURRENT FILING DATE: 2003-08-07
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S. 10/615,659
; PRIOR FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 13
; LENGTH: 541
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-635-977-13

Query Match 99.6%; Score 2843; DB 4; Length 541;

Best Local Similarity 99.6%; Pred. No. 2.5e-228;

Matches 539; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MASSILKVVVSHQSCSRSPRQREAGSDLSRRQDAENAEAKLGLPGQLVDIAC 60

Db 1 MASSILKVVVSHQSCSRSPRQREAGSDLSRRQDAENAEAKLGLPGQLVDIAC 60

Qy 61 KVCQAYLGQLEHEDITSADAVEDLTAEWEDLTQYYSLVHGDAFISNSRNYFSQCQAL 120

Db 61 KVCQAYLGQLEHEDITSADAVEDLTAEWEDLTQYYSLVHGDAFISNSRNYFSQCQAL 120

Qy 121 LNRITSVNPQTDIDGLRNIWIKPAKSRGRDIVCMRVEILELAADHPLSRDNKVV 180

Db 121 LNRITSVNPQTDIDGLRNIWIKPAKSRGRDIVCMRVEILELAADHPLSRDNKVV 180

Qy 181 QKIETPLLICDTKFDIRQWFLVTDWNPFTIWFYKESYLRFTQSFSLDKLDSAIHLCNN 240

Db 181 QKIETPLLICDTKFDIRQWFLVTDWNPFTIWFYKESYLRFTQSFSLDKLDSAIHLCNN 240

Qy 241 AVQKYLKNDVGRSPLLPANHMTSTRFQEYLQGRGAVGWSVIYPSMKKAIAHAMKVAQ 300

Db 241 AVQKYLKNDVGRSPLLPANHMTSTRFQEYLQGRGAVGWSVIYPSMKKAIAHAMKVAQ 300

Qy 301 DHVEPRKNSFELYGADFVLGRDPRPWLIEINSSPTMHPSTPTAQLCAQVQEDTIKVAVD 360

Db 301 DHVEPRKNSFELYGADFVLGRDPRPWLIEINSSPTMHPSTPTAQLCAQVQEDTIKVAVD 360

Qy 361 RSCDIGNFELLWRQVPVPPPPSGSDLCVAGSVRRARRQVLPVCLKASALLDAQPLK 420

Db 361 RSCDIGNFELLWRQVPVPPPPSGSDLCVAGSVRRARRQVLPVCLKASALLDAQPLK 420

Qy 421 ARGPSAMPDPAQGPSALQRLDGLKEKGLPLALLAPLRGAESGGAAQPTRTKAAGV 480

Db 421 ARGPSAMPDPAQGPSALQRLDGLKEKGLPLALLAPLRGAESGGAAQPTRTKAAGV 480

Qy 481 ELPACPCRHVDSQAPNTGVPVAQPAKSWDPNQLNAHPLPVLRLKLTAEGLRPPPGKG 540

Db 481 ELPACPCRHVDSQAPNTGVPVAQPAKSWDPNQLNAHPLPVLRLKLTAEGLRPPPGKG 540

Qy 541 S 541

Db 541 S 541

RESULT 5

US-10-615-659-14

; Sequence 14, Application US/10615659
; Publication No. US20040157234A1

GENERAL INFORMATION:

APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; CURRENT FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 14
; LENGTH: 293
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-615-659-14

Query Match 54.7%; Score 1560; DB 4; Length 293;

Best Local Similarity 100.0%; Pred. No. 1.5e-121;

Matches 293; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 73 EDITSADAVEDLTAEWEDLTQYYSLVHGDAFISNSRNYFSQCQALLNRITSVNPQTD 132

Db 1 EDITSADAVEDLTAEWEDLTQYYSLVHGDAFISNSRNYFSQCQALLNRITSVNPQTD 60

Qy 133 IDGLRNIWIKPAKSRGRDIVCMRVEILELAADHPLSRDNKVVQKIETPLLICD 192

Db 61 IDGLRNIWIKPAKSRGRDIVCMRVEILELAADHPLSRDNKVVQKIETPLLICD 120

Qy 193 TKFDIRQWFLVTDWNPFTIWFYKESYLRFTQSFSLDKLDSAIHLCNNAVQKYLKNDVGR 252

Db 121 TKFDIRQWFLVTDWNPFTIWFYKESYLRFTQSFSLDKLDSAIHLCNNAVQKYLKNDVGR 180

Qy 253 SPLLPANHMTSTRFQEYLQGRGAVGWSVIYPSMKKAIAHAMKVAQDHVEPRKNSFEL 312

Db 181 SPLLPANHMTSTRFQEYLQGRGAVGWSVIYPSMKKAIAHAMKVAQDHVEPRKNSFEL 240

Qy 313 YGADFVLGRDPRPWLIEINSSPTMHPSTPTAQLCAQVQEDTIKVAVDRSCDI 365

Db 241 YGADFVLGRDPRPWLIEINSSPTMHPSTPTAQLCAQVQEDTIKVAVDRSCDI 293

RESULT 6

US-10-635-977-14

; Sequence 14, Application US/10635977
; Publication No. US20040171131A1

GENERAL INFORMATION:

APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635,977
; CURRENT FILING DATE: 2003-08-07
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S. 10/615,659
; PRIOR FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 14
; LENGTH: 293
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-635-977-14

Query Match

Best Local Similarity 54.7%; Score 1560; DB 4; Length 293;

Matches 293; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 73 EDITSADAVEDLTAEWEDLTQYYSLVHGDAFISNSRNYFSQCQALLNRITSVNPQTD 132

Db 1 EDITSDADAVEDTEAEWEDLTQYYSLVHGDAFISNSRNYFSQCALNRTISVNPQTD 60
QY 133 IDGLRNIWIKPAKSRGRDIVCMRVEEILELAADHPLSRDNKVVQKIETPLLICD 192
Db 61 IDGLRNIWIKPAKSRGRDIVCMRVEEILELAADHPLSRDNKVVQKIETPLLICD 120
QY 193 TKFDIRQWFLVTDWNPNTIWFYKESYLRSTQRFSLDKLDSAIHLCNNAVQKYLKNDVGR 252
Db 121 TKFDIRQWFLVTDWNPNTIWFYKESYLRSTQRFSLDKLDSAIHLCNNAVQKYLKNDVGR 180
QY 253 SP LLPAHNMWMTSTRFOEYLQRQGRGAVGWSVIYPSMKKAIAHAMKVAQDHVEPRKNSFEL 312
Db 181 SP LLPAHNMWMTSTRFOEYLQRQGRGAVGWSVIYPSMKKAIAHAMKVAQDHVEPRKNSFEL 240
QY 313 YGADFVLGRDRFRLWLEINSSPTMHPSTPVTQALCAQVQEDTIKVAVDKSCDI 365
Db 241 YGADFVLGRDRFRLWLEINSSPTMHPSTPVTQALCAQVQEDTIKVAVDKSCDI 293

RESULT 7
US-10-615-659-4
; Sequence 4, Application US/10615659
; Publication No. US20040157234A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; CURRENT FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4
; LENGTH: 242
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-615-659-4

Query Match 45.5%; Score 1299; DB 4; Length 242;
Best Local Similarity 100.0%; Pred. No. 7e-100;
Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 133 IDGLRNIWIKPAKSRGRDIVCMRVEEILELAADHPLSRDNKVVQKIETPLLICD 192
Db 1 IDGLRNIWIKPAKSRGRDIVCMRVEEILELAADHPLSRDNKVVQKIETPLLICD 60
QY 193 TKFDIRQWFLVTDWNPNTIWFYKESYLRSTQRFSLDKLDSAIHLCNNAVQKYLKNDVGR 252
Db 61 TKFDIRQWFLVTDWNPNTIWFYKESYLRSTQRFSLDKLDSAIHLCNNAVQKYLKNDVGR 120
QY 253 SP LLPAHNMWMTSTRFOEYLQRQGRGAVGWSVIYPSMKKAIAHAMKVAQDHVEPRKNSFEL 312
Db 121 SP LLPAHNMWMTSTRFOEYLQRQGRGAVGWSVIYPSMKKAIAHAMKVAQDHVEPRKNSFEL 180
QY 313 YGADFVLGRDRFRLWLEINSSPTMHPSTPVTQALCAQVQEDTIKVAVDKSCDI 372
Db 181 YGADFVLGRDRFRLWLEINSSPTMHPSTPVTQALCAQVQEDTIKVAVDKSCDI 240
QY 373 RQ 374
Db 241 RQ 242

RESULT 8
US-10-635-977-4
; Sequence 4, Application US/10635977
; Publication No. US20040171131A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN

; TITLE OF INVENTION: TYROSINE-LIGASE-LIKE PROTEIN, BGS42
; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635,977
; CURRENT FILING DATE: 2003-08-07
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S. 10/615,659
; PRIOR FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4
; LENGTH: 242
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-635-977-4

Query Match 45.5%; Score 1299; DB 4; Length 242;
Best Local Similarity 100.0%; Pred. No. 7e-100;
Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 133 IDGLRNIWIKPAKSRGRDIVCMRVEEILELAADHPLSRDNKVVQKIETPLLICD 192
Db 1 IDGLRNIWIKPAKSRGRDIVCMRVEEILELAADHPLSRDNKVVQKIETPLLICD 60
QY 193 TKFDIRQWFLVTDWNPNTIWFYKESYLRSTQRFSLDKLDSAIHLCNNAVQKYLKNDVGR 252
Db 61 TKFDIRQWFLVTDWNPNTIWFYKESYLRSTQRFSLDKLDSAIHLCNNAVQKYLKNDVGR 120
QY 253 SP LLPAHNMWMTSTRFOEYLQRQGRGAVGWSVIYPSMKKAIAHAMKVAQDHVEPRKNSFEL 312
Db 121 SP LLPAHNMWMTSTRFOEYLQRQGRGAVGWSVIYPSMKKAIAHAMKVAQDHVEPRKNSFEL 180
QY 313 YGADFVLGRDRFRLWLEINSSPTMHPSTPVTQALCAQVQEDTIKVAVDKSCDI 372
Db 181 YGADFVLGRDRFRLWLEINSSPTMHPSTPVTQALCAQVQEDTIKVAVDKSCDI 240
QY 373 RQ 374
Db 241 RQ 242

RESULT 9
US-10-275-595A-5
; Sequence 5, Application US/10275595A
; Publication No. US20040078804A1
; GENERAL INFORMATION:
; APPLICANT: YUE, Henry
; APPLICANT: TANG, Y. Tom
; APPLICANT: AU-YOUNG, Janice
; APPLICANT: LU, Dyung Aina M.
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: HILLMAN, Jennifer L.
; APPLICANT: AZIMZAI, Yalda
; APPLICANT: LAU, Preeti
; APPLICANT: YAO, Monique G.
; APPLICANT: BURFORD, Neil
; APPLICANT: KEARNEY, Liam
; APPLICANT: POLICKY, Jennifer L.
; TITLE OF INVENTION: CYTOSKELETON-ASSOCIATED PROTEINS
; FILE REFERENCE: PE-0772 USN
; CURRENT APPLICATION NUMBER: US/10/275,595A
; CURRENT FILING DATE: 2003-06-13
; PRIOR APPLICATION NUMBER: US 60/201,960
; PRIOR FILING DATE: 2000-05-05
; PRIOR APPLICATION NUMBER: US 60/202,729
; PRIOR FILING DATE: 2000-05-08
; PRIOR APPLICATION NUMBER: US 60/209,705
; PRIOR FILING DATE: 2000-06-05
; PRIOR APPLICATION NUMBER: US 60/210,149
; PRIOR FILING DATE: 2000-06-07
; PRIOR APPLICATION NUMBER: US 60/213,215

Db 114 NPVWMDGKWWVQKQYIERPLLIIFTGKFDLRQWFLVTDWNPVLTWFWYRDSYIRFSTQPFSL 173
QY 229 DKLDSAIHCNNAVKYKLNQVGRSPLLPAHNNWTSRFOEYVLRQGRGAVNGSVIYPSM 288
Db 174 KNLDSNVHLNNSIQKHLNSCHRHPLPPDNMWSQRFOAHLQEMGAPNASTIIVPGM 233
QY 289 KKAIAHAMKVAQDHVPRKNSPELYGADFVLGRDRPWLIEINSSPTMHPSTPVTQAOLCA 348
Db 234 KDAAVHALQTSQDTVQCRKASFELYGADFVFGSDFOFWLIEINASTMAPSTAVTARLCA 293
QY 349 QVQEDTIKAVV----DRSCDIGNFELLWRQPVVEPPFPFSGSDLCVAGSVRR 396
Db 294 GVQADTLRVVIDEXLDNRCDTGAFELIYKQPAVEVQYVGIRLLVBSFTIKK 345
RESULT 12
US-10-615-659-7
; Sequence 7, Application US/10615659
; Publication No. US20040157234A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; PRIOR FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 7
; LENGTH: 352
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-615-659-7

Query Match 33.5%; Score 956.5; DB 4; Length 352;
Best Local Similarity 54.2%; Pred. No. 4.8e-71;
Matches 179; Conservative 55; Mismatches 91; Indels 5; Gaps 2;
QY 74 DIDSADAVEDLTAEWEDLTQOYVSLVHGDAFISNSRNVFSQCQALLNRITSVNPQTDI 133
Db 2 DIDKDLAEPYLTPEGWSLFLQRYQVHGEAELRHLDTVQRCEDILQQLQAVVPQIDM 61
QY 134 DGLRNIWIKPAKSRGRDIVCMDRVVEILELAADHPLSRDNKNVWVQKVIETPLLICDT 193
Db 62 EGRDNIWIKPGAKSRGRGIMCMHLEMLKLVNGNPVVMKDGKWWVQKVIETPLLIFFT 121
QY 194 KFDIRQWFLVTDWNPVTWYFQESYLRFSQRLDLSAIHLNNAVKYKLNQVGRS 253
Db 122 KFDLRQWFLVTDWNPVTWYFQESYLRFSQRLDLSAIHLNNAVKYKLNQVGRS 181
QY 254 PLIPAHNMWTSRFOEYVLRQGRGAVNGSVIYPSMKAIAHAMKVAQDHVPRKNSPELY 313
Db 182 PLLPDDNNWSQRFOAHLQEMGAPNASTIIVPGMKDAVHALQTSQDTVQCRKASFELY 241
QY 314 GADFVLGRDRPWLIEINSSPTMHPSTPVTQAOLCAQVQEDTIKAVV----DRSCDIGNFE 369
Db 242 GADFVFGEDFQPWLIENASPTMAPSTAVTARLCAGVQADTLRVVDRMLDRNCDTGAFE 301
QY 370 LLMRQPVVEPPFPFSGSDLCVAGV-SVRRAR 398
Db 302 LIYKQPVTTSPASTPRPSCLLPMYSSTRAR 331

RESULT 13
US-10-635-977-7
; Sequence 7, Application US/10635977
; Publication No. US20040171131A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; PRIOR FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 7
; LENGTH: 352
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-635-977-7

; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635,977
; PRIOR FILING DATE: 2003-08-07
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S. 10/615,659
; PRIOR FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 7
; LENGTH: 352
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-635-977-7
Query Match 33.5%; Score 956.5; DB 4; Length 352;
Best Local Similarity 54.2%; Pred. No. 4.8e-71;
Matches 179; Conservative 55; Mismatches 91; Indels 5; Gaps 2;
QY 74 DIDSADAVEDLTAEWEDLTQOYVSLVHGDAFISNSRNVFSQCQALLNRITSVNPQTDI 133
Db 2 DIDKDLAEPYLTPEGWSLFLQRYQVHGEAELRHLDTVQRCEDILQQLQAVVPQIDM 61
QY 134 DGLRNIWIKPAKSRGRDIVCMDRVVEILELAADHPLSRDNKNVWVQKVIETPLLICDT 193
Db 62 EGRDNIWIKPGAKSRGRGIMCMHLEMLKLVNGNPVVMKDGKWWVQKVIETPLLIFFT 121
QY 194 KFDIRQWFLVTDWNPVTWYFQESYLRFSQRLDLSAIHLNNAVKYKLNQVGRS 253
Db 122 KFDLRQWFLVTDWNPVTWYFQESYLRFSQRLDLSAIHLNNAVKYKLNQVGRS 181
QY 254 PLIPAHNMWTSRFOEYVLRQGRGAVNGSVIYPSMKAIAHAMKVAQDHVPRKNSPELY 313
Db 182 PLLPDDNNWSQRFOAHLQEMGAPNASTIIVPGMKDAVHALQTSQDTVQCRKASFELY 241
QY 314 GADFVLGRDRPWLIEINSSPTMHPSTPVTQAOLCAQVQEDTIKAVV----DRSCDIGNFE 369
Db 242 GADFVFGEDFQPWLIENASPTMAPSTAVTARLCAGVQADTLRVVDRMLDRNCDTGAFE 301
QY 370 LLMRQPVVEPPFPFSGSDLCVAGV-SVRRAR 398
Db 302 LIYKQPVTTSPASTPRPSCLLPMYSSTRAR 331

RESULT 14
US-10-756-149-5267
; Sequence 5267, Application US/10756149
; Publication No. US20050181375A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSIS OF METASTATIC CANCER, COMPOSITIONS AN
; TITLE OF INVENTION: METHODS OF SCREENING, FOR MODULATORS OF METASTATIC CANCER
; FILE REFERENCE: file
; CURRENT APPLICATION NUMBER: US/10/756,149
; CURRENT FILING DATE: 2004-01-12
; NUMBER OF SEQ ID NOS: 5818
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 5267
; LENGTH: 352
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-756-149-5267
Query Match 33.5%; Score 956.5; DB 5; Length 352;
Best Local Similarity 54.2%; Pred. No. 4.8e-71;
Matches 179; Conservative 55; Mismatches 91; Indels 5; Gaps 2;
QY 74 DIDSADAVEDLTAEWEDLTQOYVSLVHGDAFISNSRNVFSQCQALLNRITSVNPQTDI 133
Db 2 DIDKDLAEPYLTPEGWSLFLQRYQVHGEAELRHLDTVQRCEDILQQLQAVVPQIDM 61
QY 134 DGLRNIWIKPAKSRGRDIVCMDRVVEILELAADHPLSRDNKNVWVQKVIETPLLICDT 193

Db 62 EGDRIWIIVKPAKSGRIGMCMHLEMLKLVNGNPVVMKDGKVVQYIERPLLIFFT 121
Qy 194 KFDIRQWFLVTDNPLTIWFYKESYLRFSQFSLDKLDSAIHLCCNNAVQKYLKNDVGRS 253
Db 122 KFDLRQWFLVTDNPLTIWVYRDSYIRFSTQFSLKNDNSVHLCCNNSIQKHLNSCHRH 181
Qy 254 PLLPAHNNWTSTRFOEYIQRQGRGAVGWSVIYPSMKKAIAHAMKVAQDHVPRKNSFELY 313
Db 182 PLLPPDNMWSQRFQAHLEMGAPNAWSTIIVPGMKDAVIHALQTSQDTVQCRKASFELY 241
Qy 314 GADVLGRDPRFWLIEINSPTHPSPTVTAQLCAQVQEDTIKVAV----DRSCDIGNFE 369
Db 242 GADVFVGEDFQFWLIEINASPTMAPSTAVTARLCAGVQADTLRVVIDRMLDRNCDTGAFE 301
Qy 370 LLWRQVVEPPFPFSGSDLCVAV-SVRRAR 398
Db 302 LIYKQPVTTSPASTPRPSCLLPMYSDTRAR 331

RESULT 15

US-10-108-260A-4209
; Sequence 4209, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20040005560A1e1 full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 4209
; LENGTH: 326
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-108-260A-4209

Query Match 33.1%; Score 944.5; DB 4; Length 326;
Best Local Similarity 54.3%; Pred. No. 4.3e-70;
Matches 176; Conservative 55; Mismatches 86; Indels 7; Gaps 2;
Qy 74 DIDTSADAVEDLTEAEWEDLTQYYSLVHGDAFISNSRNYFSQCOALLNRITSVNPQTDI 133
Db 2 DIDKLEAPLYLTPEGWSLFLQRYQVWVHEGAELRHLDTQVQRCEDILQQLQAVVPQIDM 61
Qy 134 DGLRNWIKPAKSGRDIIVCMDRVETLELAADHPILSRDNKWKVQYIETPLIICDT 193
Db 62 EGDRIWIIVKPAKSGRIGMCMHLEMLKLVNGNPVVMKDGKVVQYIERPLLIFFT 121
Qy 194 KFDIRQWFLVTDNPLTIWFYKESYLRFSQFSLDKLDSAIHLCCNNAVQKYLKNDVGRS 253
Db 122 KFDLRQWFLVTDNPLTIWVYRDSYIRFSTQFSLKNDNSVHLCCNNSIQKHLNSCHRH 181
Qy 254 PLLPAHNNWTSTRFOEYIQRQGRGAVGWSVIYPSMKKAIAHAMKVAQDHVPRKNSFELY 313
Db 182 PLLPPDNMWSQRFQAHLEMGAPNAWSTIIVPGMKDAVIHALQTSQDTVQCRKASFELY 241
Qy 314 GADVLGRDPRFWLIEINSPTHPSPTVTAQLCAQVQEDTIKVAV----DRSCDIGNFE 369
Db 242 GADVFVGEDFQFWLIEINASPTMAPSTAVTARLCAGVQADTLRVVIDRMLDRNCDTGAFE 301
Qy 370 LLWRQVVEPPFPFSGSDLCVA 390
Db 302 LIYKQPVTTSPASTPRPSCLLPMYSDTRAR 331

Search completed: April 4, 2006, 12:57:59
Job time : 78.9322 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2006, 12:46:06 ; Search time 21.1171 Seconds
(without alignments)
2118.072 Million cell updates/sec

Title: US-10-635-977-2
Perfect score: 2854
Sequence: 1 MASSILKVVVSHQSCRSR.....LRGLKTAEGALRPPPGKGS 541

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/1/iaa/5 COMB.pap.*
2: /cgn2_6/ptodata/1/iaa/6 COMB.pap.*
3: /cgn2_6/ptodata/1/iaa/H COMB.pap.*
4: /cgn2_6/ptodata/1/iaa/pCTUS COMB.pap.*
5: /cgn2_6/ptodata/1/iaa/RE COMB.pap.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	280.5	9.8	439	2	US-10-104-047-3658
2	271.5	9.5	592	2	US-10-104-047-3371
3	212	7.4	561	2	US-09-270-767-46703
4	183	6.4	404	2	US-10-104-047-3621
5	154.5	5.4	260	2	US-09-270-767-62411
6	154.5	5.4	507	2	US-09-270-767-46787
7	151	5.3	259	2	US-09-248-796A-14603
8	144.5	5.1	753	2	US-09-538-092-31
9	126	4.4	305	2	US-09-270-767-46577
10	126	4.4	2482	2	US-09-252-991A-16967
11	125.5	4.4	802	2	US-09-823-240A-2
12	123	4.3	884	6	5208144-8
13	120	4.2	542	2	US-09-949-016-6778
14	120	4.2	554	2	US-09-949-016-11687
15	120	4.2	950	2	US-09-252-991A-25927
16	114.5	4.0	1319	1	US-08-290-731C-2
17	114.5	4.0	1336	1	US-08-290-731C-6
18	113.5	4.0	2074	2	US-09-491-356C-9
19	111	3.9	806	2	US-09-902-540-14992
20	110.5	3.9	560	2	US-09-252-991A-25999
21	107.5	3.8	535	2	US-09-252-991A-17140
22	107	3.7	419	2	US-10-002-344A-237
23	107	3.7	1183	2	US-09-134-001C-3530
24	106.5	3.7	328	2	US-09-252-991A-17729
25	106.5	3.7	574	2	US-09-270-767-45812
26	106	3.7	679	2	US-09-252-991A-18857
27	106	3.7	780	2	US-09-252-991A-32892

28	105.5	3.7	542	2	US-09-252-991A-32139	Sequence 32139, A
29	105.5	3.7	1706	2	US-09-252-991A-31760	Sequence 31760, A
30	105	3.7	416	2	US-09-252-991A-32660	Sequence 32660, A
31	105	3.7	561	2	US-09-252-991A-23080	Sequence 23080, A
32	105	3.7	781	2	US-09-949-016-9170	Sequence 9170, Ap
33	104.5	3.7	582	2	US-08-906-863-3	Sequence 3, Appli
34	104.5	3.7	582	2	US-09-129-668-3	Sequence 18696, A
35	104	3.6	406	2	US-09-252-991A-18696	Sequence 8369, Ap
36	104	3.6	955	2	US-09-949-016-8369	Sequence 21872, A
37	103.5	3.6	248	2	US-09-252-991A-21872	Sequence 8, Appli
38	103.5	3.6	798	2	US-09-150-460B-8	Sequence 7, Appli
39	103.5	3.6	875	2	US-09-150-460B-7	Sequence 6, Appli
40	103.5	3.6	1290	2	US-09-150-460B-6	Sequence 20033, A
41	103	3.6	420	2	US-09-252-991A-20033	Sequence 32630, A
42	103	3.6	515	2	US-09-252-991A-32630	Sequence 159, App
43	103	3.6	691	2	US-09-758-759-159	Sequence 3402, Ap
44	103	3.6	710	2	US-10-104-047-3402	Sequence 12305, A
45	102.5	3.6	398	2	US-09-902-540-12305	

ALIGNMENTS

RESULT 1

US-10-104-047-3658
; Sequence 3658, Application US/10104047
; Patent No. 6943241
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. 6943241el full length cdna
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 3658
; LENGTH: 439
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-104-047-3658

Query Match	9.8%	Score 280.5;	DB 2;	Length 439;
Best Local Similarity	25.0%	Pred. No. 3e-18;		
Matches	101;	Conservative 65;	Mismatches 135;	Indels 103; Gaps 18;
QY	82	VEDLTEARWE	-----DLT-----	QOYSLVHGDAFISNRNYFSQC 117
DB	65	VEVKDEGEWDFYWCDSWLRNFHDTYMDHVRISHFRNHYELTRKNYVVKLFRKQL 124		
QY	118	QALLNRITSVN	---PQT-----	DIQGLN---IWIIPKPAKSRGRDIVCMRDVE 160
DB	125	EREAGLEAAKCDFFKTFEMPCEYHLEFVEFRKNPGITWIMKPVARSQKGFILFRRLK 184		
QY	161	EILELAADHPLSRDNK	-----WVQKYETETPLLICDTKFDIROWFLVTDWNPITWF 213	
DB	185	DIVDM-RKDTSSDDQDDIPVENYVAQRYIENFVLCGRKFDLEVVYLVMS	-----V 236	
QY	214	YKESVLRPSTQFSLDKLDSAIHLCCNNAVKYLKNDVGRSPLLPANH	-----MMTSTRFQ 268	
DB	237	FAECLLWSGHR	-----QDVHLTNVAQK-----	TSPDYHPKKGCKWTLQFR 279
QY	269	EYL-QROGRGAWGVSIVPSMKKAIAHAMKVAQDHVEPRKNSFELYGADFVLCGRDFRWL 327		
DB	280	QYLASKHGPEAV	--ETLFRDIDNTFVKSLSQVQKVIISDKHCFELYGYDILIDQDLKMWL 337	
QY	328	IEINSSPTMHPSTPVTALCAQVQEDITKVA	-----VDRSCDIGNFELLWRQPVVE--- 378	
DB	338	LEVNASPLTASSQEDYELKTCLELTLHVVDMEARLTGCRKRVGGFGLMNDGVPVSRRE 397		
QY	379	-PPFPGSGLCVAGSV	-----RRARRQVLPCNL-----KASA 411	

Db 398 GAPDLSGMGNFVINTHLGCYVNDKQKQLRQLF--CSLOVQKASS 439

RESULT 2

US-10-104-047-3371
; Sequence 3371, Application US/10104047
; Patent No. 6943241
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. 6943241el full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3371
; LENGTH: 592
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-104-047-3371

Query Match 9.5%; Score 271.5; DB 2; Length 592;
Best Local Similarity 29.8%; Pred. No. 3.6e-17;
Matches 72; Conservative 43; Mismatches 80; Indels 47; Gaps 8;
Qy 113 YFSQCQALLNRITTSVNPQTDIDGLRNIIWIKPAKSRGRDIVCMRVVEILELAADHPL 172
Db 195 YFQERQMLGK-----HSYWKCPAELSRGRLIFSDPKFI----- 232
Qy 173 SRNKNVWVQKIETPLICTKFDIROWFLVTDWNPITWIFYKESYLRFTQSFSLDKL- 231
Db 233 -FDMYIVQKYSINPLLIIGRYKCDLRIYVCVTGPKLTIYVQSEGLVRFATEKFDLSNLQ 291
Qy 232 DSATHLCNNVQK-----YLKNDVGRSPLLPAHNMWTSRFOYLORQGRGAVW---GS 282
Db 292 NNVAHLNINSKSGASYEKIKEVIGHG-----CKWTLRFRFSYLRK-----WDVDDL 339
Qy 283 VIYPSMKKATAHAMKVAQDHVEPRKNSFELYGADFVLGRDPRPWLIRINSPP--TMEHPST 340
Db 340 LLWKIHRMWLITLITAPSVPFAANCFFELFGFDILIDNNLKPWLLEVNYSPALTLDCST 399
Qy 341 PV 342
Db 400 DV 401

RESULT 3

US-09-270-767-46703
; Sequence 46703, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 46703
; LENGTH: 561
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-46703

Query Match 7.4%; Score 212; DB 2; Length 561;
Best Local Similarity 29.2%; Pred. No. 1.9e-11;
Matches 57; Conservative 37; Mismatches 75; Indels 26; Gaps 6;
Qy 179 WVKYIETPLICTKFDIROWFLVTDWNPITWIFYKESYLRFTQSFSLDKL-DSAI 235
Db 1 VVSKYIVDPICIDGHKCDLRYVLTFSFDPLLIYIEEGIVRLATVKYDRHADNLNPPCM 60

Qy 236 HLCNNAVOKYLN-----DVGRSPLLPAHNMWTSRFOEYLORQG---RGAYWG- 281
Db 61 HLCNYSINKYHSYIRSSDAQDEVG-----HKWTLISALLRHUKLQSCDTRQLMLNI 112
Qy 282 -SVIYPSMKKATAHAMKVAQDHVEPRKNSFELYGADFVLGRDPRPWLIRINSPTMHPST 340
Db 113 EDLIKAVLACQSIISACRMFYPNGNCCFELYGFDLIDNAHKPWLLEINLSPSMGVDS 172
Qy 341 PVTAQLCAQVQEDTI 355
Db 173 PLDTKVKKCLMADLL 187
RESULT 4
US-10-104-047-3621
; Sequence 3621, Application US/10104047
; Patent No. 6943241
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. 6943241el full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3621
; LENGTH: 404
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-104-047-3621

Query Match 6.4%; Score 183; DB 2; Length 404;

Best Local Similarity 25.5%; Pred. No. 7.3e-09;
Matches 82; Conservative 38; Mismatches 118; Indels 84; Gaps 17;
Qy 6 LKVVSHQSCSRSRKPRQREEAAGSSDLSSRQDAENAEAK-----LRGLPGQLVDIA 59
Db 119 LKW-----CEVKSRSYSGSFRE--GEQLLYQLPNKKLLTKIGLLSTLRG-RARAMSRA 169
Qy 60 CKVQAVYLGLEHEDIDTSADAVEDLTEAEWEDLTQYYSILVHGDATISNRNYFSQCA 119
Db 170 SKVPGGVQARLEK---DAAAPALEDL---PW-----TSPGYLRQRV 205
Qy 120 LLNRITSVNPQT---DIDGLR-----NIWIKPAKSRGRDIVCMRVVEILELA 166
Db 206 L--RMBEFPFETVRLDLKHEREAFFTLFDETIQWICKPTASNOCKGIFLLRNQBEVAALQ 263
Qy 167 A-----ADHPLSRDNKW-----VVKYIETPLICTKFDIROWFLVTDWNPITWIFYK 215
Db 264 AKTRSMEDDPIHKKTPFRGPQARVWQYIQNPLLVGKRKFDVRSYLIACCTPYMI-FFG 322
Qy 216 ESYLRFTQSFSLDKLDSAJHLCNNAVOKYLNKNDVGRSP---LLPAHNMWTSRFOEYL- 271
Db 323 HGVARLUTLSLDHSSDLGGHLTNQFMQK-----KSPLYMLLKERTVSMHEHLNRYIS 375
Qy 272 ---QROGRGAVGMSVIYPSMK 289
Db 376 DTFWKARGLAKDW---VFTTLK 394
RESULT 5
US-09-270-767-62411
; Sequence 62411, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17

```

; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 62411
; LENGTH: 260
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-62411

```

Query Match 5.4%; Score 154.5; DB 2; Length 260;
Best Local Similarity 31.7%; Pred. No. 2e-06;
Matches 40; Conservative 18; Mismatches 43; Indels 25; Gaps 3;

```

RESULT 6
US-09-270-767-46787
; Sequence 46787, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and
; FILE REFERENCE: File Reference: 7326-0
; CURRENT APPLICATION NUMBER: US/09270,
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 46787
; LENGTH: 507
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-46787

```

Query Match 5.4%; Score 154.5; DB 2; Length 507;
Best Local Similarity 31.7%; Pred. No. 6e-06;
Matches 40; Conservative 18; Mismatches 43; Indels 25; Gaps 3;

```

RESULT 7
US-09-248-796A-14603
; Sequence 14603, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AM
; FILE OF INVENTION: FOR DIAGNOSTICS AND AM
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13

```

```

; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 14603
; LENGTH: 259
; TYPE: PR1
; ORGANISM: Candida albicans
US-09-248-796A-14603

```

Query Match	5.3%;	Score 151;	DB 2;	Length 259;
Best local Similarity	22.6%;	Pred. No. 4.4e-06;		
Matches 53;	Conservative 50;	Mismatches 81;	Indels 50;	Gaps 9

Qy	132	DIDGLRWIWKPAKSRGRDI	---	VCMDRVVEIL	-----	ELAADHP	-----	L 172
		: : : : : :	: :	: : : : : :		:		
Db	20	EINKEEKLWLKPSMSDKGIR	FTK	LDLQLOE	IFNSFEENDEE	DEGVDEENGIL		79
Qy	173	SRNKWVVKYIETPLLICDTE	DIQWEL	---	VDNNPLTIWPKSYLRFSTQ	RFSL	---	228
		: : : : : :	: :	: : : : : :		:		
Db	80	SQLEHFIVQBYKSNPLLL	--	SKYDHKKFHLRTYVCV	GDCLKVFVYKXNLT	TLTFAGEPYKLP		137
Qy	229	----	DKLDSAIHL	CNNAAQVKYLKND	VGRSPLLPAHNNMTW	STRFQEYLQRCGRGAV	WGSVI 284	
		: : : : : :	: :	: : : : : :		:		
Db	138	GDEDEVSLAGHLTNTCLQ	E----	NEDPLV	-----	VFFKQLQGLADN	KNDKIV 180	
Qy	285	YPSK----	KKAIAHAMKVAQDH	VEPRKNSFELYGAD	FVLGRDRFPWLI	ETNSSP 334		
		: : : : : :	: :	: : : : : :		:		
Db	181	FEQICDITKELFKAATS	VDKNNFQIPINNA	ETIEIFGVDELVNS	SDFSVNLLEVNSYP 234			

```

RESULT 8
US-09-538-092-31
; Sequence 31, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Glot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538, 092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuratPatSeqFormatter Version 0.9
; SEQ ID NO 31
; LENGTH: 753
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number YBR094W
US-09-538-092-31

```

Query Match	5.1%	Score 144.5;	DB 2;	Length 753;
Best Local Similarity	21.8%;	Pred. No. 0.00011;		
Matches	57;	Conservative 44;	Mismatches 105;	Indels 56; Gaps 9
QY	140	WIIPKAASGRGRDVCMDRVEEI-	-----LELAADHP-----	171
Db	476	WIVKPSMSDKGQIRVFKTIEDLQAFSDPDDSEAESGNDDDDADDVNGEFMNNKVN		535
QY	172	LSGRDNKVVQKYIETELLIC----	DTKPIRQWFLVTDNPLTTIWFYKESYLFSTQRP--	226
Db	536	ISQLRHFIQEVLTNPILLASMDNRKFHRCY--	VVCRGDLQVFVYDRMLALFAAKFPVP	593
QY	227	-----SLDKLDSATHLCNNAVKYLKNDVGRSPLLPAHMMWTSTRFQ--	EYLQROGRG	277
Db	594	LDPYAVSVTDLKDLECHLTNTCLQS--KKGDKSSVL-----	EFDSIEETPNRKS	642

```
QY 278 AVWGSVIYPSMKKAAIAHAMKVAQDHVEPRKNSPELYGADFLVGRDPRFWLIEINSSPTMH 337
Db 643 NI-KEQHSITNDVFLAANNVNRNFOPLPNAPETGVDFLIDNSYEVKLEINAPDFK 701
QY 338 PSTPVTQAQLCAQVQEDTIKVAV 359
Db 702 QTGKDLKNLIDELFDDTVKVCV 723

RESULT 9
US-09-270-767-46577
; Sequence 46577, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 46577
; LENGTH: 305
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; US-09-270-767-46577

Query Match 4.4%; Score 126; DB 2; Length 305;
Best Local Similarity 23.6%; Pred. No. 0.0015;
Matches 42; Conservative 36; Mismatches 62; Indels 38; Gaps 8;

QY 180 VQKIETPLICTKDFDIROFLVTDWNPPLTIWYFYESYL-RFSTORFSLDKLDSAIHLC 238
Db 1 VQSFVORPYLVGHKFDIGYVVVITSVNLPLRYITGDVLFRCYKVPKYPHFDNAE----- 55
QY 239 NNAVQYKLVKNDVGRSPLLP-----HNW---TSTRFOEVLORQGR--GAVWGSVI 284
Db 56 ---VDKVIYGD---DYLFTWEPVSLKYYNRFEGGSMRTVFEAYVRDQGDPAQIWPQVE 108
QY 285 Y-----PSMKKAAIAHAMKVAQDHVEPRKNSPELYGADFLVGRDPRFWLIEINSSPTM 336
Db 109 HIVRTTIAAEKDIVNLIRSYRTH-----NFFDLMRFDLFIDEDLKVFLMEANWSPNL 161

RESULT 10
US-09-252-991A-16967
; Sequence 16967, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 16967
; LENGTH: 2482
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (1766)
; OTHER INFORMATION: Identity of amino acid at the above locations are unknown.
; US-09-252-991A-16967

Query Match 4.4%; Score 126; DB 2; Length 2482;
Best Local Similarity 19.7%; Pred. No. 0.044;
Matches 122; Conservative 79; Mismatches 207; Indels 210; Gaps 29;
```

```
QY 71 EHEDIDT-----SADAVEDLTAEWEDLTQQYVYLHVGDAFISNSRNYFSQCOALLN 122
Db 308 DHETVDAERARLAGPRDRDNRSVVVGALCBELVRIKDSL---DLFVSRDRGHSELDALLA 364
QY 123 RITSVNPQTDIDG-----LNNIWIIPAAKSRG--DIVCMDRVBEILEAA----- 167
Db 365 PLKQIADTLAVLGFQGPQPRKVIQDLDVIHALAAGRRPSDAIILMDVAGALLYVEATLAGM 424
QY 168 ---ADHPLSRDNKVVQKIET-PLLICDTKDFIRO-----WFLVTDWNPPLTIWYKE- 216
Db 425 AGPGDERNSEESRLPTTIDVAQHOLVIKEARNGLEQAKDAIEFFIASQNMHEHLARVP 484
QY 217 -SVLRESTQRFSLDKLDSAIHLNNAVQYKLVKNDVGRSPLLPANNWTS-----TRF 267
Db 485 LTQVRGGLAMIPLERAAITLETCTNRYIQEQL---LARK---AVPDWQSLDITLADAITSV 537
QY 268 QEYLQQRGAVGWSVIYPSMKKAAIAHAMKVAQDHVEPRKNSPELYGADFLVGRDPRFWL 327
Db 538 EYTLERL-----SEDHASQSDLI---LDVAEDSLAN-----LGYTLKP-- 572
QY 328 IEINS-----SPTMHSTPTVAQLCAQ-----VOEDTIKVA 358
Db 573 --NSSAPAEPLSGSPAAIESPAAEPEPEAVVEVAETAEPQPPADTAPAEAREDAPOLA 629
QY 359 VDSCDIG-----NFELLWRQPV---VEPPP-----F 382
Db 630 SDNWTITGEVADPAGPSLDLALDLDDSAEVPFALPEVVEESGQPSQTPAPARSLDDF 689
QY 383 SGSDLCVAGSVRRARRRQVLVPCNLKASASLLDAQFLKARGPSAMPD---PAQ----- 432
Db 690 SLDEIDLGLD-----LPA-----DAAP--ASGPAALADWSLPEQWGLGDDL 729
QY 433 -GPSPALORDLGLKE-----EKGLPLALLAPLRGAESG---GAQP-----TRT 474
Db 730 AQPTAQGETIDLSLEPALSFDAPLESLEPLALEPFDGSAEQELVLDLDDLPFLDVALP 789
QY 475 KAAGKV-----ELPACPCRHVDSQAPNTGVPVQAQPAKSWDPNQLNAHPL 518
Db 790 EAEGEYSANEGSLEELDLSDDLPEVQLPEAEAEAPPAEALASAPALSLEAVMAAPV 849
QY 519 EPVLRLGLKTAEGALRPPPP 536
Db 850 QPINPPAQNPVSLPPPP 867

RESULT 11
US-09-823-240A-2
; Sequence 2, Application US/09823240A
; Patent No. 6716597
; GENERAL INFORMATION:
; APPLICANT: Frank B. Gertler
; APPLICANT: James E. Bear
; APPLICANT: Jurgen Wehlend
; APPLICANT: Joseph Loureiro
; TITLE OF INVENTION: Methods and Products for Regulating Cell
; TITLE OF INVENTION: Motility
; FILE REFERENCE: M00656.70064.US
; CURRENT APPLICATION NUMBER: US/09/823,240A
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 60/194,564
; PRIOR FILING DATE: 2000-04-03
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 802
; TYPE: PRT
; ORGANISM: Mus musculus
; US-09-823-240A-2

Query Match 4.4%; Score 125.5; DB 2; Length 802;
Best Local Similarity 24.7%; Pred. No. 0.0081;
Matches 78; Conservative 15; Mismatches 110; Indels 113; Gaps 15;
```


; Sequence 11687, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C0001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR FILING DATE: 2000-10-20
; PRIOR FILING DATE: 2000-10-20
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11687
; LENGTH: 554
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-11687

Query Match 4.2%; Score 120; DB 2; Length 554;
Best Local Similarity 22.5%; Pred. No. 0.015;
Matches 75; Conservative 38; Mismatches 124; Indels 96; Gaps 18;
QY 254 PLLPAHNMWTSRFRQYELQQRGAVGVSIVPSMKKAIHAAMKVAQDHVEPRKNSFELY 313
DB 245 PLLPGNDV-----GRSS-YGAM---QVQVFDYAVIVLSHAVSPARSYPNR 287
QY 314 GADFVLGR-----DFRPLIEINSSPTMHPSTPTVTAOLCAQVOEDTIKVAVDKSCD 364
DB 288 DAESTLGRLLIKVTQEVIDYRWIKWKGSKA-HPSPGMDSRI--KIKE---RIATCNGEQ 341
QY 365 IGFNELLWRQPVVEPP-----PFGSDLCVAGSVRRARR-----QVLPVCNLKA- 409
DB 342 TQN-----REF--ESPYQORLTLSSPQLLSGSSASSVSSLSGSDVSDTTPCTTPSV 394
QY 410 -SASLLDAQPLKARGSPAMPDPAQPPSPALQDGLKEKGLPLALLAPLRGAA-----E 464
DB 395 YQFSLQAPAPLMAGLPTALPMPG--GKQPTTSRTLIMTNNQTFPIPPPTLGVAPVPCR 453
QY 465 SGGAAPTRTKAAGKVELPACPCRHVDQAPNTGVPVQAQPAKSWDPNQLNAHPLPVLRG 524
DB 454 QAGVEGTASLKAHVHMSSPAIP-----SASP-----PLSSP-----HLYHKQHG 494
QY 525 LK-----TAEGALRPPPGKG 540
DB 495 MKLSMKGSHGHTQGGYSSVSGGVPPVGNRG 527

RESULT 15
US-09-252-991A-25927
; Sequence 25927, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 25927
; LENGTH: 950
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-25927

Query Match 4.2%; Score 120; DB 2; Length 950;
Best Local Similarity 25.1%; Pred. No. 0.036;
Matches 53; Conservative 17; Mismatches 63; Indels 78; Gaps 9;
QY 361 RSCDIGNPELLWRQPVVEPPF---SGSDLCVAGVSVR-----RARRQVLPVCNLKA 409
DB 728 RTADVS-----RQPLRQFPFALLQRRGDVLDASLAPRPPGFGGAARARLLLSPPACRRGR 781
QY 410 SASLLDAQP-----LKARGPSAM-----PDPAQG-----PFSPALQRLGLK 446
DB 782 ATGLGQRQPGRGRLSRCFGAATAAGPGGTAAHARGGGDGLRRRHPPAPAPRPGPLA 841
QY 447 EKGLPLALLAPLRGAAESGGAAPTRTKAAGKVELPACPCRHVDQAPNTGVPVQAQPAK 506
DB 842 P-----APAGGAADFGAAAPGRWPA-----PAARRQR 869
QY 507 SWDPNQLNAHPLPVLRLGLKTAEGALRPPPG 537
DB 870 TWRTRRRS-----LRTLERPAGVPPAPP 894

Search completed: April 4, 2006, 12:46:55
Job time : 22.1171 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.
OM protein - protein search, using sw model
Run on: April 4, 2006, 12:43:06 ; Search time 81.9545 Seconds
(without alignments)
4657.352 Million cell updates/sec
Title: US-10-635-977-2
Perfect score: 2854
Sequence: 1 MASSILKVVVSHQSCSRSSR.....LRGLKTABGALRPPPGKGKS 541

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 2166443 seqs, 705528306 residues
Total number of hits satisfying chosen parameters: 2166443
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database : UniProt.05.80.*
1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES					Description	
Result No.	Score	Query Match %	Length	ID		
1	2741	96.0	747	Q5JZ84	human sapien	
2	1753.5	61.4	781	Q8CON7	mus musculus	
3	1726.5	60.5	518	Q8COV2	mus musculus	
4	1052	36.9	704	Q8BV51	mus musculus	
5	1043.5	36.6	744	Q9H876	human sapien	
6	981.5	34.4	434	Q8NDN8	human sapien	
7	957.5	33.5	352	Q6AWA3	human sapien	
8	956.5	33.5	352	1 TLLJ3	human sapien	
9	941.5	33.0	534	Q4KMS8	human sapien	
10	941.5	33.0	534	Q4RY08	tetraodon n	
11	834	29.2	266	Q922T0	mus musculus	
12	807.5	28.3	261	Q58CT2	bov taurus	
13	746.5	26.2	572	Q7Q156	anopheles g	
14	722.5	25.3	992	Q9VM91	drosohila	
15	584.5	20.5	501	Q7PMD3	anopheles g	
16	536.5	18.8	719	Q9VM92	drosohila	
17	535.5	18.8	756	Q5BHY1	drosohila	
18	524.5	18.4	281	Q6ZU95	human sapien	
19	524	18.4	375	Q4V8C1	rattus norv	
20	442.5	15.5	331	Q5TN29	anopheles g	
21	385	13.5	101	Q96GG8	human sapien	
22	345.5	12.1	461	Q641W7	rattus norv	
23	345	12.1	464	Q9D570	mus musculus	
24	344.5	12.1	523	Q54TU1	dictyosteli	
25	335	11.7	1075	Q7QT13	giardia lam	
26	331.5	11.6	794	Q6BFH6	paramesim	
27	328	11.5	496	Q961I9	drosohila	
28	328	11.5	496	Q9VX74	drosohila	
29	317	11.1	341	Q9VX47	human sapien	
30	315.5	11.1	618	Q4R7H0	macaca fasc	
31	314	11.0	403	2 Q7QZC8	giardia lam	

32	311	10.9	673	2	Q6ZVT0	human sapien
33	307	10.8	559	2	Q8CFV5	mus musculus
34	307	10.8	600	2	Q5T2W4	human sapien
35	307	10.8	1333	2	Q8CHB8	mus musculus
36	303.5	10.6	1048	2	Q9P1V5	human sapien
37	303.5	10.6	1226	2	Q9UFZ4	human sapien
38	303.5	10.6	1277	2	Q6EMB2	human sapien
39	303.5	10.6	1299	2	Q5R978	pongo pygma
40	301	10.5	438	2	Q8C125	mus musculus
41	301	10.5	1339	2	Q5F498	chick galli
42	299.5	10.5	423	1	TTLL1	human sapien
43	298.5	10.5	423	1	TTLL1	mouse
44	298.5	10.5	423	2	Q543S4	mus musculus
45	298.5	10.5	423	2	Q5PPI9	rattus norv

ALIGNMENTS

RESULT 1
Q5JZ84 HUMAN PRELIMINARY; PRT; 747 AA.
AC Q5JZ84;
DT 10-MAY-2005 (Tremblrel. 30, Created)
DT 10-MAY-2005 (Tremblrel. 30, Last sequence update)
DE OTTHUMP0000028514 (Fragment).
GN ORFNames=RP3-355C18.2-002;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]_TaxID=9606;
RP NUCLEOTIDE SEQUENCE.
RA Cobley V.;
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL022327; CAI42686.1; -; Genomic DNA.
DR GO; GO:0004835; F:tubulin-tyrosine ligase activity; IEA.
DR GO; GO:0006464; P:protein modification; IEA.
DR InterPro; IPR004344; Tub_tyr_lygase.
DR Pfam; PF03133; TTL; 1.
FT NON TER 1
SQ SEQUENCE 747 AA; 83854 MW; F90948E159BBE589 CRC64;
Query Match 96.0%; Score 2741; DB 2; Length 747;
Best Local Similarity 94.0%; Pred. No. 1.1e-181;
Matches 529; Conservative 4; Mismatches 6; Indels 24; Gaps 4;

Qy	1	MASSILKVVVSHQSCSRSSR	KPRDQREAGSSDLSSRQDAENAEAKRLGPGQLVDIAC	60
Db	187	MASSILKVVVSHQSCSRSSR	KPRDQREAGSSDLSSRQDAENAEAKRLGPGQLVDIAC	246
Qy	61	KVCQAYLGQLEHEDITSAD	VEDLTEAWEDLTQQYVSLVH---GDAPF-----	107
Db	247	KVCQAYLGQLEHEDITSAD	VEDLTEAWEDLTQQYVSLVQVPLGSSIVLCIFKIQVM	306
Qy	108	-----SNSRNVFSCOALL	NRITSVNPQTDIDGLRNIIWIKPAKSRGR----	158
Db	307	MSFPPPTARD--ROCAL	LNRIITSVNPQTDIDGLRNIIWIKPAKSRGRGESP	364
Qy	159	VEEILEAAADPHLSRDN	KNKVVQKYIETPLICTDKFDIROWFLVTDNPLTI	218
Db	365	VEEILEAAADPHLSRDN	KNKVVQKYIETPLICTDKFDIROWFLVTDNPLTI	424
Qy	219	LRFTQRFSLDKLSA	IHLCNNVQKYLKNDVGRSPLLPAAHMMTSTF	278
Db	425	LRFTQRFSLDKLSA	IHLCNNVQKYLKNDVGRSPLLPAAHMMTSTF	484
Qy	279	VWGSVIYPSMKKATA	HAHKVAQDHVEPRKNSFELYGADFVLGRDPR	338
Db	485	VWGSVIYPSMKKATA	HAHKVAQDHVEPRKNSFELYGADFVLGRDPR	544

```

QY 339 STPVTATLCAQVQEDTIKVAVDSCDIGNFELLWROPVVEPPFSGSDLCVAGSVRRAR 398
Db 545 STPVTATLCAQVQEDTIKVAVDSCDIGNFELLWROPVVEPPFSGSDLCVAGSVRRAR 604
QY 399 RQVLPCVNLKASASLLDAQPLKARGPSAMPDPQAQPPSPALQRLDGLKEEKGJPLALLAP 458
Db 605 RQVLPCVNLKASASLLDAQPLKARGPSAMPDPQAQPPSPALQRLDGLKEEKGJPLALLAP 664
QY 459 LRGAASGGGAQPTRTYKAGKVLPACPCRHVDSPAPNTGVPVQAQPAKSWDPNQLNAHPL 518
Db 665 LRGAASGGGAQPTRTYKAGKVLPACPCRHVDSPAPNTGVPVQAQPAKSWDPNQLNAHPL 724
QY 519 EPVLRGLKTAEGALRPPPGKGS 541
Db 725 EPVLRGLKTAEGALRPPPGKGS 747

RESULT 2
O8CON7_MOUSE
AC O8CON7_MOUSE PRELIMINARY; PRT; 781 AA.
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Mus musculus adult male testis cDNA, RIKEN full-length enriched
DE library, clone:4933401B17 product:hypothetical Tubulin-tyrosine ligase
DE containing protein, full insert sequence.
GN Names:170001901Rik;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridea; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Testis; DOI=10.1016/S0076-6879(99)03004-9;
RX MEDLINE=99279233; PubMed=10349636;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RN Meth. Enzymol. 303:19-44(1999).
[2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Testis; DOI=10.1039/35055500;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staehli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Schenwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
[3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
[4]

```

```

RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Testis; DOI=10.1101/gr.145100;
RX MEDLINE=20499374; PubMed=11042159;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
[5]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Testis; DOI=10.1101/gr.152600;
RX MEDLINE=20530913; PubMed=11076861;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitzunai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
[6]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shingawa A., Shiroki T., Sogabe Y., Tegami M.,
RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR ENBL; AK030151; BAC26811.1; -; mRNA.
DR Ensembl; ENSMUSG00000022388; Mus musculus.
DR MGI; MGI:1922902; 170001901Rik.
DR GO; GO:0016874; F:ligase activity; IEA.
DR GO; GO:0004835; F:tubulin-tyrosine ligase activity; IEA.
DR GO; GO:0006464; P:protein modification; IEA.
DR InterPro; IPR000276; GPCR_Rhodpsn.
DR InterPro; IPR004344; Tub_Tyr_ligase.
DR Pfam; PF031133; TTL; 1.
DR PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; UNKNOWN_1.
KW Hypothetical protein; Ligase.
SQ SEQUENCE 781 AA; 89398 MW; 9E98793C3351C3DE CRC64;

Query Match 61.4%; Score 1753.5; DB 2; Length 781;
Best Local Similarity 63.0%; Pred. No. 4.9e-113;
Matches 345; Conservative 69; Mismatches 113; Indels 21; Gaps 8;

QY 1 MASSILKVVVSHOS-CSSRSRSPRQREBAGSSDLSSRODAENAEAKRLGLPQLVDIA 59
Db 206 VAASILKVVVHLQNYC---SKVGKSKKEAKNSDPKDPENPDLKLSLSQVVDTA 262

QY 60 CKVCQAYLGQLEHEDITSADAVEDTEAWEDLTQYYSLVHGDAFISNRNFSQCOA 119
Db 263 CKVCQAYLGQLEHEDIDVSEASTEALSEEWNDLTQYYLLVHGNSITDSKSYFAQCOA 322

QY 120 LLMRITSVNPQTDIGLRNIIWKPAKSRGRDIVCMRVREELILEAAADHPLSRDNKWV 179
Db 323 LLSKISSVNPQTDIGIRNIIWKPAKSRGRDIVCMRVREELISLVAADSQTTKDNKWV 382

QY 180 VQKYIETPLICDTKFDIROWFLVTDWNPPLTIWFYKESYLFSTQRFSLDKLSAIHLN 239
Db 383 VQKYIETPLIYDTKFDIROWFLVTDWNPPLTIWFYKESYLFSTQRFSLDKLSAIHLN 442

QY 240 NAVQYKLVNDVGRSPLLPALNMMTSTRFQBYLQQRQGVAGVWGSVIYPSMKKAIHAKMVA 299
Db 443 NSIQRRLLKNDKERSPLIPCHNMMTSTRFQBYLQQRQGVAGVWGSVIYPSMKRAVTNAMRVA 502

```



```

QY 360 --DRSCDIGNFELLWRQPVVPPFSGSDLCVAGSVRRARRQVLPVCNLIKASASLLDAQ 417
Db 472 RLDSCDGTGAFELIYKQPAVEVQYVIGIRLLVEGTTIK-----510
QY 418 PLKARGPSAMPDPAQGPSPALQDRLGLKEEKGPLALLAPLRGAESGGAOQPTRTKAA 477
Db 511 -----PVPVGHRTGVRS-----LPHLLTQGGESKDSGSPH-RGA 548
QY 478 GKVELPACPCRHHVDSQAPNTGVVPAQP-----AKSWDPNQLNAHPLEPVLR-- 523
Db 549 SRKNARAESLEHTEKPEPAVASVSGKGKAPHPFPLSLHSAWLPSPVRHPRQGRVLRQ 608
QY 524 -----GLKTAEGALRPPGCK 539
Db 609 HDQLVGSKALSTTGKALMTLPTAK 632

RESULT 5
Q9H876 HUMAN
ID Q9H876_HUMAN PRELIMINARY; PRT; 744 AA.
AC Q9H876;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE Hypothetical protein FLJ13898.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Thyroid gland;
RX PubMed=14702039; DOI=10.1038/ng1285;
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA Sekine M., Obayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
RA Yamamoto J.-I., Saito K., Kawai Y., Isono Y., Nakamura Y.,
RA Nagahori K., Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M.,
RA Shiratori A., Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H.,
RA Sugawara M., Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E.,
RA Omura Y., Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M.,
RA Yamazaki M., Ninomiya K., Ishibashi T., Yamashita H., Murakawa K.,
RA Fujimori K., Tanai H., Kimata M., Watanabe M., Hiraoaka S., Chiba Y.,
RA Ishida S., Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T.,
RA Kusano J., Kanehori K., Takahashi-Fujii A., Hara H., Tanase T.-O.,
RA Nomura Y., Togiani S., Komai F., Hara R., Takeuchi K., Arita M.,
RA Imose N., Musashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,
RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,
RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
RA Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T., Nakagawa K.,
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;
RT "Complete sequencing and characterization of 21,243 full-length human
RT cDNAs.";
RL Nat. Genet. 36:40-45 (2004).
DR EMBL; AK03960; BAB14741.1; -; mRNA.
DR Ensembl; ENSG00000156983; Homo sapiens.
DR GO; GO:0016874; F:ligase activity; IEA.
DR GO; GO:0004835; F:tubulin-tyrosine ligase activity; IEA.
DR GO; GO:0006464; P:protein modification; IEA.
DR InterPro; IPR004344; Tub_tyr_lygase.
DR Pfam; PF03133; TTL; 1.
KW Ligase.

```

```

FT NON_TER 744 744
SQ SEQUENCE 744 AA; 84683 MW; DF661753E4AFFODF CRC64;

Query Match 36.6%; Score 1043.5; DB 2; Length 744;
Best Local Similarity 40.5%; Pred. No. 1e-63;
Matches 234; Conservative 85; Mismatches 192; Indels 67; Gaps 12;

QY 2 ASSILKWVYSHOSCSRSSRSKPRDQREAGSSDLSRQDAENAEAKLGLPGLQVLDIACK 61
Db 151 ARNVKLKLVKSE-----WKSYPITQAVEEASGDQPKQEKQNPVL---VSPEFVDEALC 201
QY 62 VQAYILQLEHEDIDTSADAVEDLIEAWEDLTQYVSLVHGDAFTISNSRNYFSQOALL 121
Db 202 ACEEYLSNLAHMDIDKLEAPLYLTPEGWSLFLQRYVYVHEGAELRHLDLTQVQRCEDIL 261
QY 122 NRITSVNPOTDIDGLRNIIWIKPAKSRGRDITVCMRVEEILELAAAHPLSLSDNKKVVQ 181
Db 262 QQLQAVVFOIDMEGRNIIWIVPGAKSRGRGIMCMHLEMLKLVNGNPVVMKDGKVVQ 321
QY 182 KYIETPLLCIDTKFDIQWFLVTDWNLPTIWFYKESYLFSTQRFSLDKLSDAIHLNNA 241
Db 322 KYIERPLLIFFTGLRQWFLVTDWNLPTVWFYRDSYIRFSTQPFSLKNLDSNVHLNNS 381
QY 242 VQYILKNDVGRSPLLPANHWITSTRPQYVLRQGRGAVGVSITYPSMKKAIHAAMKVAQD 301
Db 382 IQKHLNSCHRHPLLPDPNMWSSQRFQHLQEMGAPNAWSTIIVPGMKDAVIHALQTSQD 441
QY 302 HVEPRKNSPELYGADPVLGRDPRPMLIEINSSPTWHPSTPVTQAOLCAQVOEDTIKVA-- 359
Db 442 TVQCRKASFELYGADPVFGEDFQWLLIEINASTTMAPSTAVTARLCAGVQADLRLVIDR 501
QY 360 --DRSCDIGNFELLWRQPVVPPFSGSDLCVAGSVRRARRQVLPVCNLIKASASLLDAQ 417
Db 502 MLDNRCDTGAFLIYKQPAVEVQYVIGIRLLVEGTTIK-----PMACHRRMGVRPAV 555
QY 418 PLKARGPS-----AMPDPAQGPSPALQDRLGL-----KKEKG 450
Db 556 PLLTQRGSGEARHHPFSLHTKAQ-LPSPHVLRHQGVLRQHSKLVTGKALSTTGKALRT 614
QY 451 LPLA-----LLAPLRGAESGGAOQPTRKA-----AGKVELPACPCRHHVDSQ--AP-NTG 498
Db 615 LPTAKVFTSLPPLNDFKVPASILKPKRKAPALLCLRGPOLEVPCCCLCKPLKSEQFLAPVGRS 674
QY 499 VPAQPAKSWDPNQNLNAHPLEPVLRGLKTAEGALRPPP 536
Db 675 REPANSRPDCDKPRAEACPMKEL-----SPLKPLP 704

RESULT 6
Q9NDN8 HUMAN
ID Q9NDN8_HUMAN PRELIMINARY; PRT; 434 AA.
AC Q9NDN8;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Hypothetical protein DKFZp586B0320.
GN Homo sapiens (human)
OS Homo sapiens (human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Uterus;
RA Wambutt R., Heubner D., Mewes H.W., Cassenhuber J., Wiemann S.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP TISSUE=PCR rescued clones;
RC MEDLINE=2388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner C.M., Schuler G.D.,

```


RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hate S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalusz D.E.,
 RA Schnurch A., Schein J.E., Jones S.J.M., Marra M.A.,
 RA "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [3]
 RP NUCLEOTIDE SEQUENCE.
 RC TISSUE=PCR rescued clones;
 RG NIH MGC Project;
 RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL033393; CAD38794.1; -; mRNA.
 DR EMBL; BC098298; AA98298.1; -; mRNA.
 DR GO; GO:0004835; F:tubulin-tyrosine ligase activity; IEA.
 DR GO; GO:0006464; P:protein modification; IEA.
 DR InterPro; IPR004344; Tub_tyr_ligase.
 DR Pfam; PF01133; TTL; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 434 AA; 94933 MW; 9E79E6CA08651C1A1 CRC64;

Query Match 34.4%; Score 981.5; DB 2; Length 434;
 Best Local Similarity 46.7%; Pred. No. 9.9e-60;
 Matches 194; Conservative 64; Mismatches 106; Indels 51; Gaps 5;

QY 74 DIDSADAVEDLTEAEWEDLTQYYSLVHGDAFISNSRNYFSQOALLNRITSVNPOTDI 133
 DB 2 DIDKLEAPLYLTPEGWSLFLQRYQVHGAEALRLHDTQVQRCEDILQQLQAVVPQIDM 61
 QY 134 DGLRNIIWIKPAKSRGRDIVCMRDVEILELAADHPLSRDNKVVQVQKIETPLLI 193
 DB 62 EGRNIIWIKPGAKSRGRGIMCNDHLEMLKLVNGPNVVMKDGKVVQVQKIETPLLI 121
 QY 194 KFDIROWFLVDNPLTIWPKESYLFSTQSFSLDKLSAIHLCNNAVOKYLNKDVGRS 253
 DB 122 KFDLRQWFLVDNPLTVWFVYRDSYIRFSTQSFSLKNDLSVHLNNSIOKHLNCHRH 181
 QY 254 PLLPAHNMWTSRFOEVLQRCRGAVGWSVIYPSMKKAIHAAMKVAADHVEPRKNSPELY 313
 DB 182 PLLPPDNWSSQRFQALHLOENGAPNAWSTIIVPGMKDAVIHALQTSQDTVQCRKASPELY 241
 QY 314 GADPVLGRDPRPWLIEINSSPTMHPSTPTVTAQLCAQVQEDTIKAVV----DRSCDIGNFE 369
 DB 242 GADVFGEDEQPWLEINASPTWAPSTAVTARLCAGVQADTLRVVIDRLDRNCDTGAFE 301
 QY 370 LLMRQPVVEPPFPFSGDLGVAGSVRRARRQVLPVCNKLKASASLLDAQPLKARGPSAMD 429
 DB 302 LIYQPAVEVPQYVIRLLVEGFTIKK-----PMAM-- 332
 QY 430 PAQGPSPALORDGLGKEEKLPLALLAPLGAESGGAQPTTKAAGKVELPA 484
 DB 333 -----CHRRMGVR-----PAVPLLTQRGSGEARHHPFLSLTKA-----QLPS 369

RESULT 7
 Q6AWA3_HUMAN
 ID Q6AWA3_HUMAN PRELIMINARY; PRT; 352 AA.
 AC Q6AWA3;
 DT 25-OCT-2004 (TrEMBLrel. 28, Created)
 DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
 DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
 DE Hypothetical protein DKFZp686D076.

GN Name=DKFZp686D076;
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae;
 OC Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC TISSUE=Cervix;
 RG The German cDNA Consortium;
 RA Ansoorge W., Krieger S., Regier T., Rittmuller C., Schwager B.,
 RA Mewes H.W., Weil B., Amid C., Osanger A., Fobo G., Han M., Wiemann S.;
 RA Submitted (SEP-2004) to the EMBL/GenBank/DBJ databases.
 RL EMBL; BX648175; CAH10554.1; -; mRNA.
 DR GO; GO:0004835; F:tubulin-tyrosine ligase activity; IEA.
 DR GO; GO:0006464; P:protein modification; IEA.
 DR InterPro; IPR004344; Tub_tyr_ligase.
 DR Pfam; PF01133; TTL; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 352 AA; 40257 MW; 49FD8E9CAE1CB20D CRC64;

Query Match 33.5%; Score 957.5; DB 2; Length 352;
 Best Local Similarity 54.2%; Pred. No. 3.5e-58;
 Matches 179; Conservative 55; Mismatches 91; Indels 5; Gaps 2;

QY 74 DIDSADAVEDLTEAEWEDLTQYYSLVHGDAFISNSRNYFSQOALLNRITSVNPOTDI 133
 DB 2 DIDKLEAPLYLTPEGWSLFLQRYQVHGAEALRLHDTQVQRCEDILQQLQAVVPQIDM 61
 QY 134 DGLRNIIWIKPAKSRGRDIVCMRDVEILELAADHPLSRDNKVVQVQKIETPLLI 193
 DB 62 EGRNIIWIKPGAKSRGRGIMCNDHLEMLKLVNGPNVVMKDGKVVQVQKIETPLLI 121
 QY 194 KFDIROWFLVDNPLTIWPKESYLFSTQSFSLDKLSAIHLCNNAVOKYLNKDVGRS 253
 DB 122 KFDLRQWFLVDNPLTVWFVYRDSYIRFSTQSFSLKNDLSVHLNNSIOKHLNCHRH 181
 QY 254 PLLPAHNMWTSRFOEVLQRCRGAVGWSVIYPSMKKAIHAAMKVAADHVEPRKNSPELY 313
 DB 182 PLLPPDNWSSQRFQALHLOENGAPNAWSTIIVPGMKDAVIHALQTSQDTVQCRKASPELY 241
 QY 314 GADPVLGRDPRPWLIEINSSPTMHPSTPTVTAQLCAQVQEDTIKAVV----DRSCDIGNFE 369
 DB 242 GADVFGEDEQPWLEINASPTWAPSTAVTARLCAGVQADTLRVVIDRLDRNCDTGAFE 301
 QY 370 LLMRQPVVEPPFPFSGDLGVAGSVRRARRQVLPVCNKLKASASLLDAQPLKARGPSAMD 398
 DB 302 LIYQPAVEVPQYVIRLLVEGFTIKK-----PMAM-- 331

RESULT 8
 TTL3_HUMAN
 ID TTL3_HUMAN STANDARD; PRT; 352 AA.
 AC Q9Y4R7; Q9UI99;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 13-SEP-2003 (Rel. 48, Last annotation update)
 DE Tubulin tyrosine ligase-like protein 3 (HOTT1).
 GN Name=TTL3; ORFNames=PRO0207;
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae;
 OC Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
 RC TISSUE=Testis;
 RG The German cDNA consortium;
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA] OF 61-352.
 RC TISSUE=Fetal liver;
 RA Yu Y., Zhang C., Luo L., Ouyang S., Zhang S., Li W., Wu J., Zhou S.,

[illegible]

```
OS Tetraodon nigroviridis (Green puffer).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;
OC Tetraodontidae; Tetraodontidae; Tetraodon.
OX NCBI_TaxID=99883;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Jaillon O., Aury J.M., Brunet F., Petit J.L., Stange-Thomann N.,
RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segures B.,
RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
RA Anthouard V., Jubin C., Castelli V., Katinka M., Vacherie B.,
RA Blemont C., Skalli Z., Cattolico L., Boulain J., De Berardinis V.,
RA Cruaud C., Duprat S., Brottier P., Coutanceau J.P., Gouzy J.,
RA Parra G., Lardier G., Chapple C., McKernan K.J., McEwan P., Bosak S.,
RA Kellis M., Wolff J.N., Guigo R., Zody M.C., Mesirov J.,
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
RA Wincker P., Lander E.S., Weissenbach J., Roest Crolius H.;
RA "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
RT the early vertebrate proto-karyotype.";
RL Nature 431:946-957 (2004).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RA Genoscope; Whitehead Institute Centre for Genome Research;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR ENBL; CAE01014979; CAG06724.1; -; Genomic DNA.
SQ SEQUENCE 534 AA; 60130 MW; 62B5B8B696B662B CRC64;

Query Match 33.0%; Score 941.5; DB 2; Length 534;
Best Local Similarity 49.2%; Pred. No. 7.9e-57;
Matches 192; Conservative 59; Mismatches 118; Indels 21; Gaps 6;

QY 3 SSILKVVVSHQSCRSRSRPRDREAGSDLSRODAENAEAKLGLPGQLVDTACKV 62
DB 146 TSLLOHVVEGSSHRA-----DAKER---QLHLPEDL--SKHLFPFNAGVIATLRV 193
QY 63 QAYLGQLEHEDITSADAVEDLTEAEWEDI.TQOYSLVHGDAFISNRNYFSQOQALLN 122
DB 194 COHYLGVLEHRDIDVT---METGEVSWGAFLODYIKVHVGASVRGGGVFERCQAMLH 250
QY 123 RITSVNPQTDIDGLRNIIKPAKSRGRDVCMDRVEEILEAAADHPLSRDNKWVYVQK 182
DB 251 KIQEVCQLGTDLGLNNIIKPGAMSRGIVCKRDLDEILAVDNDKSLNKEKKWVYVQK 310
QY 183 YIETPLLCDTKFDIRQWFLVDNPLTIWFKYESYLRFSTQFSLDKLDSAIHLGNVAV 242
DB 311 YLERPLLVHGTGKFDLRQWFLVDNPLTVWFYRECYLRFSTQFSTKILDSSIHLCNNSI 370
QY 243 QYILKNDVGRSPLPAHNMWTSRFBQYLQQRGAVGWSVYPSMKKAIAHAMKVAODH 302
DB 371 QKHFEPAARNRHPAVPGDNMWSQSFRAFLHQQGRSAEWVTVFGMKAIIRALQTAQEQ 430
QY 303 VEPKNSFELYGADFVLGRDPRPWLIEINSPTHPSTPTVTAQICAOVQEDTIKVAVD 362
DB 431 VEPKNSFELYGADFVLMGRDLRPLWLEINICPTMAQSSVVTALHCLPAVQIDTLRVLD 490
QY 363 CD-----IGNFELLWRQPVPEP--PPFSGSD 386
DB 491 SDPSAHTGGFOLICQAGGSGGCPLFSGSE 520

RESULT 11
Q922T0_MOUSE PRELIMINARY; PRT; 266 AA.
ID Q922T0_MOUSE PRELIMINARY; PRT; 266 AA.
AC Q922T0;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
```

AC Q58CT2;
 DT 10-MAY-2005 (TrEMBLrel. 30, Created)
 DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
 DE 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
 DE Tubulin tyrosine ligase-like family, member 3.
 GN Names=TTLL3;
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
 OC Pecora; Bovidae; Bovinae; Bos.
 OX NCBI_TaxID=9913;
 RN [1]
 RN NUCLEOTIDE SEQUENCE.
 RP TISSUE=Pooled;
 RC MEDLINE=21180013; PubMed=11282978; DOI=10.1101/gr.170101;
 RX Smith T.P.L., Grosse W.M., Freking B.A., Roberts A.J., Stone R.T.,
 RA Casas E., Wray J.E., White J., Cho J., Fahrnerkrug S.C., Bennett G.L.,
 RA Heaton M.P., Laegreid W.W., Rohrer G.A., Chitko-Mckown C.G.,
 RA Perlea G., Holt I., Karanycheva S., Liang F., Quackenbush J.,
 RA Keele J.W.;
 RT "Sequence evaluation of four pooled-tissue normalized bovine cDNA
 RT libraries and construction of a gene index for cattle."
 RL Genome Res. 11:626-630(2001).
 RN [2]
 RN NUCLEOTIDE SEQUENCE.
 RP TISSUE=Pooled;
 RC Harhay G.P., Sonstegard T.S., Van Tassel C.P., Clawson M.L.,
 RA Heaton M.P., Keele J.W., Snelling W.M., Weidmann R.T., Smith T.P.L.;
 RT "Sequencing and analysis of Bos taurus full-length insert cDNA
 RT clones";
 RL Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BT021865; AAX46712.1; -, mRNA.
 KW Ligase.
 SQ SEQUENCE 261 AA; 29911 MW; 5F069784CA162017 CRC64;
 Query Match 28.3%; Score 807.5; DB 2; Length 261;
 Best Local Similarity 58.5%; Pred. No. 6.2e-48;
 Matches 148; Conservative 37; Mismatches 59; Indels 9; Gaps 1;
 QY 132 DIDGLRNIWIKPAKSRGRDIVCMRDVVEEILELAAADHPLSRDNKVVQKYEIPLTIC 191
 Db DMEGRNIWIKVPAKSRGRDIVCMRDVVEEILELAAADHPLSRDNKVVQKYEIPLTIC 61
 QY 192 DTKFDIQRFLVTDNPLTIWFKYSYLFSTORFSLDKLSAHLGNNAVQKYLKNDVG 251
 Db GTKFDIQRFLVTDNPLTIWFKYSYLFSTORFSLDKLSAHLGNNAVQKYLKNDVG 121
 QY 252 RSPILPAHNMWTSRFOEYLQRCGAVGWSVIYPSMKKAIAHAMKVAQDHVEPRKNSPE 311
 Db RHPLPPDNMWSQKFOAHLQETGAPNAWSTVIVFGMKAIVTHALQTSQDTVQCRKASPE 181
 QY 312 LYGADFVLGRDPRPMLIENSSPTMHPSTPTVTAQLCAQVOEDTIKAVDRSCD----- 364
 Db LYGADFVFGEDQFPMLEINASTPAPSTAVTARLCAGVQADTLRVVDWRLDRTVTRGP 241
 QY 365 --IGNFELLWRQP 375
 Db LSSSTSSLPWRCP 254
 RESULT 13
 ID Q70156 ANOGA PRELIMINARY; PRT; 572 AA.
 AC Q70156;
 DT 01-MAR-2004 (TrEMBLrel. 26, Created)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
 DE ENSANGP0000022337 (Fragment).
 GN ORFNames=ENSANG0000019848;
 OS Anopheles gambiae str. PEST.
 OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Culicidae;
 OC Anophelinae; Anopheles.

OX NCBI_TaxID=180454;
 RN [1]
 RN NUCLEOTIDE SEQUENCE.
 RC STRAIN=PEST;
 RG The Anopheles gambiae Sequence Committee;
 RT "Anopheles gambiae re-annotation."
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RN NUCLEOTIDE SEQUENCE.
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=PEST;
 RG The Anopheles gambiae Sequence Committee;
 RL Submitted (APR-2004) to the EMBL/GenBank/DBJ databases.
 CC -!- CAUTION: The sequence shown here is derived from an
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
 CC preliminary data.
 DR EMBL; AAB01008980; EAA13905.2; -, Genomic DNA.
 DR GO; GO:0004835; F.tubulin-tyrosine ligase activity; IEA.
 DR GO; GO:0006464; P.protein modification; IEA.
 DR InterPro; IPR0004344; Tub_tyr_lygase.
 DR Pfam; PF03133; TTL; 1.
 FT NON_TER 1
 FT NON_TER 572 572
 SQ SEQUENCE 572 AA; 67109 MW; 8938D9EDD5935071 CRC64;
 Query Match 26.2%; Score 746.5; DB 2; Length 572;
 Best Local Similarity 39.8%; Pred. No. 3.1e-43;
 Matches 159; Conservative 70; Mismatches 134; Indels 37; Gaps 9;
 QY 5 ILKVVVSHQSCSRSSRSPROREAGSSDLSSRODAENAEAKLRGLPGQLVDIACKVCQ 64
 Db LUKWLIERHG-----TEGGGGGL-----DAIADD--GNVPSTCITFALTACK 229
 QY 65 AYLGQLEHEDITSADAVEDLTAEWEDLTQOYSLVHGD---AFISNR-----NYFS 115
 Db EYLDYCLNDIDIEEDT--KWDHDDVFLTHYLLTHEDNRQLLKEEREADAIEHYLA 287
 QY 116 QOQALLNRTISVNPOTIDGLRNIIWIKPAKSRGRDIVCMRDVVEEILELAAADHPLSRD 175
 Db EAKSVLEQIKSHWPQVALDGYLNIWIKVPGNRCGRGIHLMNNIKI--IAMVNPPIVSK 345
 QY 176 NKWVQKYEIETPLLCIDTKFDIQRFLVTDNPLTIWFKYSYLFSTORFSLDKLSAI 235
 Db TRYIVTKYIERPLIHTKFDIQRFMITSVQPLNIWFKYSYLFSTORFSLDKLSAI 405
 QY 236 HLCNNAVQKYLKNDVGRSPILPAHNMWTSRFOEYLQRCGAVGWSVIYPSMKKAIAHA 295
 Db HLTNHAIOKKYHNAV-RDERLPHENWDCHTFOAYLRQIDKYEMMSERIPGMQKAIIGS 464
 QY 296 MKVAQDHVEPRKNSPELYGADFVLGRDPRPMLIENSSPTMHPSTPTVTAQLCAQVOEDTI 355
 Db LILACQDNMDRRPNTPELYGADFMITEDFVPMLEINSSPDLAPSTSVTARLCPCQVEDTI 524
 QY 356 KVAVDRSCD-----IGNFELLWRQPVEPPPPSGSDLCVAG 391
 Db RVVIDRRDTSNAPTSGFELIYKQVTPKTPAYMGLNLQLRG 564
 RESULT 14
 ID Q9VM91 DROME PRELIMINARY; PRT; 992 AA.
 AC Q9VM91;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE CG11323-PA.
 GN Name=CG11323; ORFNames=CG11323;
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RN NUCLEOTIDE SEQUENCE.

RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.C., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brannon R.C., Rogers J., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballwe R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Berns P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foeller C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacieb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Smith H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*.";
RL Science 287:2185-2195(2000).
[2]
RN NUCLEOTIDE SEQUENCE.
RP MEDLINE=22426065; PubMed=12537568;
RX Celinker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R.,
RA Pacieb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svirskaas R., Taber P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstein G., Scher S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT "Finishing a whole-genome shotgun: release 3 of the *Drosophila*
RT melanogaster euchromatic genome sequence.";
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
[3]
RN NUCLEOTIDE SEQUENCE.
RP MEDLINE=22426070; PubMed=12537573;
RX Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J.W., Svirskaas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celinker S.E.;
RT "The transposable elements of the *Drosophila melanogaster* euchromatin:
RT a genomes perspective.";
RL Genome Biol. 3:RESEARCH0084.1-RESEARCH0084.20(2002).
[4]
RN NUCLEOTIDE SEQUENCE.
RP MEDLINE=22426069; PubMed=12537572;
RX Miera S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Celinker S.E., de Grey A.D.N.J., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a
RT systematic review.";
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).

RN NUCLEOTIDE SEQUENCE.
RP Berkeley *Drosophila* Genome Project;
RA Celinker S., Carlson J., Wan K., Pfeiffer B., Frise E., George R.,
RA Hoskins R., Stapleton M., Pacieb J., Park S., Svirskaas R., Smith E.,
RA Yu C., Rubin G.;
RT "Drosophila melanogaster release 4 sequence.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[6]
RN NUCLEOTIDE SEQUENCE.
RP FlyBase;
RL Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL: AS003614; AAF52432.1; -; Genomic DNA.
DR Ensembl: CG11323; *Drosophila melanogaster*.
DR FlyBase: Fgn0031854; CG11323.
DR GO: GO:0004835; Fubulin-tyrosine ligase activity; IEA.
DR GO: GO:0006464; P:protein modification; IEA.
DR InterPro: IPR004344; Tub_tyr_lygase.
DR Pfam: PF03133; TTL; 1.
SQ SEQUENCE 992 AA; 112255 MW; 0FC34674C15BDAAD CRC64;
Query Match 25.3%; Score 722.5; DB 2; Length 992;
Best Local Similarity 33.3%; Pred. No. 3e-41;
Matches 175; Conservative 86; Mismatches 198; Indels 67; Gaps 12;
QY 51 LPQQLVDIAKVCQAYLQGLEHEDIDTSADAVED---LTEAEWEDLTQQYSLVHGDAFI 107
DB 326 IPYSAIDFAYKRLVEYIDSCQHNID-----FEDPKIWEHDWDAFLFQHQQLVNEDGRI 380
QY 108 SNS-----RNYFSQCQALLNRNITSVNPQTDIDGLRNIIKPKAASGRGRDIVCMRVVEI 162
DB 381 QHGGQRLSPWVKSLSLVDKMKVHPQYSLDGYQNNWIVKPAKCKRGRIILMDNLKKI 440
QY 163 LELAADHPLSRDNKWKVQKIETPLLICOTKFDIQFWELVDNPLITWFKESYLRF 222
DB 441 --LGVVNLSTASKSVVQKIERPLILFQTKFDIQFWELITNTQPLVVMFYRESYLRF 498
QY 223 TQRFSLDKLDSATHLCNNVQKYLKNDVGRSPLPAHNNMTSTRFOYLQROGRGAVWGS 282
DB 499 SQEYSLSNHHSVHLTYAIQKKYNG-KRDKLPSENMMWDCYSFOYLQIKYNNWLE 557
QY 283 VIYPSMKKATAHAMKVAQDHVEPRKNSFELYGADFVLGRDFRPLWIEINSSPTMHPSTPV 342
DB 558 RIFPGMKALVGMCLASQENMDRPNPTFELFGADFMCENFYFWLIEINSSPDLGATTSV 617
QY 343 TAQLCAQVBDTITKIVADRSCLD-----IGNFELLWRQPVVEPPFPFSGDLCAVGSV--- 394
DB 618 TARMCPQCLDGVVWIDRTDPKAEIENFELAYRQVVPPTPAYMGLNLFVKGQVLQKA 677
QY 395 -----RRARQVLPVCNLKASASLDAQPLKARGPSAMP----- 428
DB 678 NHGGHGHVYQQQKERSLATSSVTRQRSATIHPTATSIIRHAMTFNATEYMEKVMV 737
QY 429 DPAQGPSPALQDGLKKEKGLPLALLAPLRAAESGGAAQPTRTKAAGKVELPACPCR 488
DB 738 EPLSSRS-SLCSQLPKQSPSAALATATP-----SGATSSYLKQAGRSITQLLSAT 789
QY 489 H-----VDSQAPNTGVPAQPAKSWDPNQLNAHPLEPVLRLGLK 526
DB 790 HKRNTGGLSGEQVQSTALP-PKRQRSQGLPSSTNPVESTKFKF 834
RESULT 15
Q7PMD3_ANOGA
ID Q7PMD3_ANOGA PRELIMINARY; PRT; 501 AA.
AC Q7PMD3;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE ENSANGP0000022334 (Fragment).
GN ORFNames=ENSANG00000019845;
OS Anopheles gambiae str. PEST.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2006, 12:45:31 ; Search time 20.1115 Seconds
(without alignments)
2588.231 Million cell updates/sec

Title: US-10-635-977-2
Perfect score: 2854
Sequence: 1 MASSILKVVVSHQSCRSR.....LRGLKTAEGALRPPGKGKS 541

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: PIR.80.*
2: PIR1.*
3: PIR2.*
4: PIR3.*
5: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	956.5	33.5	352	T12515	hypothetical prote
2	259	9.1	379	A45443	tubulin-tyrosine 1
3	252.5	8.8	640	E88575	protein XK1128.6 [
4	252.5	8.8	680	T27699	hypothetical prote
5	209.5	7.3	1198	T20262	hypothetical prote
6	209.5	7.3	1203	C89217	protein C55A6.2 [i
7	166	5.8	662	T20343	hypothetical prote
8	161	5.6	403	T37571	tubulin-tyrosine 1
9	155.5	5.4	883	A96805	hypothetical prote
10	144.5	5.1	753	S48261	hypothetical prote
11	130	4.6	2472	E83594	still frameshift p
12	124	4.3	4660	T42737	gp330 protein proc
13	120	4.2	882	A70507	probable reductase
14	119.5	4.2	3436	S55659	tegument protein 6
15	115.5	4.0	1006	T41104	probable transcrip
16	114.5	4.0	1336	S25716	Ras guanine nucleo
17	112.5	3.9	696	T46394	hypothetical prote
18	112.5	3.9	1575	S68448	synaptotagmin, 170K
19	111	3.9	930	A25923	progesterone recep
20	109	3.8	906	A43817	transforming prote
21	108.5	3.8	741	I51657	suppressor of yeas
22	108.5	3.8	1048	T31425	C-terminal domain
23	108.5	3.8	1650	S53457	dominant autoantig
24	108.5	3.8	2176	T39188	probable U5 snRNP
25	108	3.8	940	JE0291	FB19 protein - hum
26	108	3.8	1245	G84897	hypothetical prote
27	108	3.8	3149	1 Q9BE8	BPLF1 protein - hu
28	107.5	3.8	420	T46910	hypothetical prote
29	107.5	3.8	896	B43817	transforming prote

30	107	3.7	755	2	S32103	filensin - bovine
31	107	3.7	1234	2	T00363	hypothetical prote
32	106.5	3.7	538	2	S57459	hook-containing pr
33	106.5	3.7	2870	2	H96974	cyclic beta 1-2 gl
34	105.5	3.7	517	2	E89530	protein H28G03.2 [
35	105.5	3.7	4957	2	T03455	ALR protein - huma
36	105.5	3.7	5262	2	T03454	ALR protein - huma
37	104	3.6	836	2	T32298	hypothetical prote
38	103.5	3.6	1290	2	T00018	period protein hom
39	103.5	3.6	2774	2	A43359	microtubule-associ
40	103	3.6	670	2	D86176	hypothetical prote
41	103	3.6	816	2	T15049	hypothetical prote
42	102.5	3.6	775	2	S63626	homeotic protein A
43	102.5	3.6	906	2	JCS963	stable tubule only
44	102	3.6	454	2	A84162	hypothetical prote
45	102	3.6	454	2	T08280	hypothetical prote

ALIGNMENTS

RESULT 1

T12515
hypothetical protein DKFZp434B103.1 - human
C:Species: Homo sapiens (man)
C:Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 09-Jul-2004
C:Accession: T12515
R:Wambutt, R.; Heubner, D.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, June 1999
A:Reference number: Z17524
A:Accession: T12515
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-352 <WAM>
A:Cross-references: UNIPROT:Q9Y4R7; UNIPARC:UPI00001377A6; EMBL:AL096725
A:Experimental source: adult testis; clone DKFZp434B103
C:Genetics:
A:Note: DKFZp434B103.1

Query Match 33.5%; Score 956.5; DB 2; Length 352;
Best Local Similarity 54.2%; Pred. No. 78-64; Indels 5; Gaps 2;
Matches 179; Conservative 55; Mismatches 91;

QY	74	DIDSADAVEDLTAEWEDLTQQYSLVHGDAFISNRNYFSQCALNLRITSVNPQTDI	133
DB	2	DIDKLEAPLYITPGWSLFLQRYQVHVGAEHLHLDTVQRCEDILQQLQAVVPQIDM	61
QY	134	DGLRNIIWKPAKSGRDIIVCMRVVEILELAADHPLSRDNKVVQKYIETPLICDT	193
DB	62	EGDRNIWIVKPGAKSGRGIMCMCHLEMLKLVNGNVVMKDKVQVQKYIERPLIFGT	121
QY	194	KFDIRQWELVTDWNPPLTTWYKESYLRFSTQRFSDKLSATHLCNNAVQKYLKNDVGRS	253
DB	122	KFDLRQWELVTDWNPPLTVWFYRDSYIRFSTQFSLKLNLDNSVHLNNSIQKHLNSCHRH	181
QY	254	PLLPANNMTSTRFOEYLQRCRGAVGWSVIYPSMKKAIAHAMKVAQDHVEPRKKSFEYL	313
DB	182	PLLPDNNWSSQRFQAHLEQMGAPNAWSTIIVPGMKDAVIHALQTSQDTVQCRKASFELY	241
QY	314	GADFVLGRDFRPWLIEINSSPTMHPSTPVTALCAQVQEDTIKAV-----DRSCDIGNPE	369
DB	242	GADFVFGEDFQFWLIEINASTMAPSTAVTARLCAGVQADTLRVVDRMLDRNCDTGAFE	301
QY	370	LLMQPVPVEPPFSGSDLCVAGV-SVRRAR	398
DB	302	LIYKQPVTTSPASTPRPSCLLPMYSDTRAR	331

RESULT 2

A45443
tubulin-tyrosine ligase (EC 6.3.2.25) [validated] - pig
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004

C;Accession: A45443
R;Ersfeld, K.; Wehland, J.; Plessmann, U.; Dodemont, H.; Gerke, V.; Weber, K.
J. Cell Biol. 120, 725-732, 1993
A;Title: Characterization of the tubulin-tyrosine ligase.
A;Reference number: A45443; MUID:93147125; PMID:8093886
A;Accession: A45443
A;Status: preliminary
A;Molecule type: mRNA; protein
A;Residues: 1-379 <ERS>
A;Cross-references: UNIPROT:P38160; UNIPARC:UPI000001377A9; GB:X68453; GB:S54050; NID:G21241
A;Experimental source: brain
A;Note: sequence extracted from NCBI backbone (NCBIN:123858, NCBIP:123859)
C;Keywords: ligase

Query Match 9.1%; Score 259; DB 2; Length 379;
Best Local Similarity 26.8%; Pred. No. 1.2e-11;
Matches 98; Conservative 57; Mismatches 128; Indels 82; Gaps 20;

Qy 52 PG-QLVDI---ACKVC-QAYLGLEHEDIDTSADAVEDLTEAEWEDLTQOYYSL----- 100
Db 56 PGLMLVNYRGADKLCRKASLVK-----IKTSPELAESCT---WPPESVVIPTNLKTP 108
Qy 101 -----VHGDAFISNRYSFQCALLNRTISVNPQTIDIGLRNIWIIKPAKSR 149
Db 109 VAPAQNGIHPPIHSSR--TDREFF-----LTSYNKKKE-DGEGNWIAKSSAGAK 156
Qy 150 GRDIVCMRVVEIIELEAADHPLSRDNK---WVQKVIETPLLI---CDTKFDIRQWFLVY 204
Db 157 GEGILISSEATELLDFTI-----DNQGVHVITQKYLRLPEGRHFRSRYWLVLD 208
Qy 205 DNPPLTTFWYKESLRFSTORFSLDKL-DSAHLCCNAVKYLKNDVGRSPLPAHNMWT 263
Db 209 --HQNYLYREGVLRITASPHYIDNFQDKTCHLTNHCIOKEYSKYVK---YEEGNEFM 263
Qy 264 STRFEQLORQGRGAVMGSVIYSPMKKAIHAAMKVAQDHVEPR-----KNGFELYGAD 316
Db 264 FEEFNQVLT-----SALNITLESILLQIKHIIRSCLLSVEPAISRHLVPQSOLFQFGD 318
Qy 317 FVLGRDPRPMLIENSSPTMHPSTPTVAQLCAQVQEDTIKVAVDRSCDIGNFELLWQPV 376
Db 319 FMVDEDLKWLLIEYNGAPAC--AOKLYAELC---QGIVDIAIASVFPFPPDAE----QQQ 368
Qy 377 VEPPP 381
Db 369 QQPPP 373

RESULT 3
E88575
C;Species: Caenorhabditis elegans
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
C;Accession: E88575
R;anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biological
A;Reference number: A75000; MUID:95069613; PMID:9851916
A;Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/projects/C_elegans/
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A;Accession: E88575
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-640 <STO>
A;Cross-references: UNIPROT:Q09647; UNIPARC:UPI0000017A077; GB:chr_III; PIDN:CAA87425.1;
C;Genetics:
A;Gene: ZK1128.6
A;Map position: 3

Db	283	ETDASRHV-IVKPPASAGRTGISVTRKPKDFPTATL-----VAQHYIERPLTIN	331
Qy	192	DTYFEDIHQWFLVTDWNPITLTFWKESVLRSTQRFSLDK---LDSAIHLCHNAVQKYLKN	248
Db	332	RAKFDRLYAYVPTFEFLRYIYDQGLVRFASVPYSHSVTISNKYMHLTNTYSINKLAEB	391
Qy	249	D-VGRSPL--LPAHNMTWTSRF-----QEYLQRQGRGAVMGSVIYPSMKKAIAHAMKVA	299
Db	392	DGVANKRPVKWTLHLW--EHFDEMGVDRKIQRE-----IEEVIIKAFISTEKPI	440
Qy	300	QDH-----VEPRKNSFELYGADFVLGRFRPWLIIENSPTMHPSTPTVAQLCAQVOEDTI	355
Db	441	REHMSRFLQEFCYELFGIDIILDDEYKPMLEVNITSPSLHSGTPTLDVSVKAPLAKDVL	500
Qy	356	KVAVDKSCDIGNPELLWRQPVVEPPPF---SGSDLCVAGVSVRRAROVLP-----VCNLK	408
Db	501	NLA-----GVYVEPPSFDKUSDADYSTPRNGRKNRQGLIKEASVWAAYK	544
Qy	409	ASASLLDAQPLKARGP	424
Db	545	DQLGVINDNRIFKRLTP	560
RESULT 4			
T27699			
hypothetical protein ZK1128.6 - Caenorhabditis elegans			
C/Species: Caenorhabditis elegans			
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004			
C/Accession: T27699			
R/Berks, M.			
submitted to the EMBL Data Library, January 1995			
A/Reference number: Z20407			
A/Accession: T27699			
A/Status: preliminary; translated from GB/EMBL/DBJ			
A/Molecule type: DNA			
A/Residues: 1-680 <WIL>			
A/Cross-references: UNIPROT:Q09647; UNIPARC:UPI000017A076; EMBL:Z47357; PIDN:C			
A/Experimental source: clone ZK1128			
C/Genetics:			
A/Gene: CESP:ZK1128.6			
A/Map position: 3			
A/Introns: 114/1; 208/3; 242/3; 334/3; 362/2; 405/2; 424/2; 550/2; 633/2			
Query Match			
Beat Local Similarity 8.8%; Score 252.5; DB 2; Length 680;			
Matches 81; Conservative 25.6%; Pred. No. 8.3e-11;			
Matches 81; Conservative 56; Mismatches 118; Indels 61; Gaps 12			

Qy	132	DIDGRNIWIKPAKSRGRDVCMDRVEEILELAADHPLSRDNKWWVQVKYIETPLLIC	191
Db	323	ETDASRHLV-IVKPPASARFGTGSVTRKPKDPFTTATL-----VAQHYIERPLTTN	371
Qy	192	DTKEDIQOWELVTDWNPLTIWFYKESYLRFSTQRESLDK---LDSAIHLCCNNAVQYKLN	248
Db	372	RAKFDRLUYAVVPTFEPURVVIYDQCLVPFASVPYSHSVSTISNKVMHLTNYSinkLAE	431
Qy	249	D-VGRSPSL--LPAHNMWTSRFP-----QEYLQROGRGAVWGVSIVPSMKKAIAHAMKVA	299
Db	432	DGVANKPVPKTLHLHM--EHFDEMGVDRKEIQRE-----IEEVIKAFISTEKPI	480
Qy	300	QDH-----VEPRKNSEFLYGADPVLGRDPRFPMLEINSSPTMHPSTPTVTAQLCAQOEDTI	355
Db	481	REHMSRFLQEBFICYELFGIDIILDEDYKPMLEVNISPSLHSGTPLDVSVKAPLAKDVL	540
Qy	356	KVAVDRSCDIGNFELLWQPVVEPPPF---SGSDLCVAGSVVRRARRQVLP-----VCNLK	408
Db	541	NLA-----GVVYPSFDKLSADYSTRPNRGKRKEQLIKEASWVAAYK	584
Qy	409	ASASILLDAQPLKARGP	424
Db	585	DQLGVINDNRIFKRLTP	600

QY 132 DIDGLRNIWIIKPAAKSRGRDIVCMDRVEEILELAAADHPLSRDNKWVQKYIETPLLC 191

RESULT 8
T37571
tubulin-tyrosine ligase related protein - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C:Accession: T37571
C:R:Badcock, K.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
submitted to the EMBL Data Library, April 1996
A:Reference number: Z21726

A;Accession: T37571
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-403 <BAD>
A;Cross-references: UNIPROT:Q10438; UNIPARC:UPI000013A94B; EMBL:Z70721; PIDN:CAA94694.1;
A;Experimental source: strain 972h-; cosmid c12B10
C;Genetics:
A;Gene: SPDB:SPAC12B10.04
A;Map position: 1
A;Introns: 320/3; 348/3

Query Match 5.6%; Score 161; DB 2; Length 403;
Best Local Similarity 20.8%; Pred. No. 0.00027;
Matches 72; Conservative 65; Mismatches 127; Indels 86; Gaps 14;

QY 71 EHEIDITSADAV-----EDLTAEWEDLTQQYYSLVHGDAFISNS----- 110
DB 49 QYEDID--FDEVYKNPKTKLCCSVIRKALIRKEYLWRTVITYLAKHPDSILSKSVPEAY 106

QY 111 ---RYFSQCOALLNRITSVNPTDIDGLRNI-----WIIPKAASRGDRDIVCMRVVEI 162
DB 107 SLELDYAEFLDLSMEAYELRQLEENATKNISEKQWYILKPSMCDRAQOIRLFSTIEEL 166

QY 163 LEL-----AAADHPLSRDNK-----WVQKYIETPLLCDTKFDI 197
DB 167 QAIFDSFDDESESEAGLEEKGDITVAFNNKIVISOIRNFLVQKYSKPLLLDHRKPHI 226

QY 198 ROWFLVTWNPLTIWFYKESYLRFSTQRFSLDKLDSAI---HLCNNAVQKYLKNDVGRSP 254
DB 227 RAVLAT--GALSVYLENEMCLLARDKYKPTFPDPLFSLHSLNTCLO---GDNVEQSS 281

QY 255 LLPAHNMWTSRFOEYLQRGGAVMGVSVIYPSMKXAIHAAMKVAQD---HVEPRKNSF 310
DB 282 I---RDFWNTS-----IENK-----DQIFKSIILNIIGDVFEAAATTOGIFQPLENCF 326

QY 311 ELYGADFVLGRDPRPMLIENSSFTWHPSTPTVAQLCAQVOEDTIKVAVD 360
DB 327 EIFGVDFLVDCEQVYLLEVNS----YPDPKOTGKNLSNIENLFSASVE 372

RESULT 9
A96805
hypotheical protein T5M16.14 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C;Accession: A96805
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Feder-spiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzial,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: A96805
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-883 <STO>
A;Cross-references: UNIPROT:Q9CAQ1; UNIPARC:UPI000000A09A3; GB:A8005173; NID:G6382502; PI
C;Genetics:
A;Gene: T5M16.14
A;Map position: 1

Query Match 5.4%; Score 155.5; DB 2; Length 883;
Best Local Similarity 25.3%; Pred. No. 0.0022;
Matches 55; Conservative 36; Mismatches 93; Indels 33; Gaps 7;

QY 134 DGLRNIWIIPKAASRGDRIVCMRVVEILELAAADHPLSRDNKWWQKYIETPLLCDT 193
DB 641 DQLNNLWILKPNWARTIDTSITNLSAIR-----MMETGPKICQKIIEHPALPKGN 693

QY 194 KFDIROWFLVTDWNPDLTIWFYKESYLRFSTQRFSLDKLDSAIHLCNNAVQKYLKNDVGRS 253
DB 694 KFDLRYVVLVRSIDPLEIYLIIFWVRLSNPNYSLEKHSFFFEYTHFTVWNY-----GRK 748

QY 254 PLLPAHNMWTSRFOEYLQROG-----RGAVMGSV-----IYPSMK---KAIHAAMK 297
DB 749 ----LNHKPTAEFVREFEQEHNDFAFHFNVTIQLSIVKWMIDIHEKVKQVIRAVFEAAA 804

QY 298 VAQDHVEPRKNSFELYGADFVLGRDPRPMLIENSSP 334
DB 805 LAHPMQSPK-SRAMYGVDMVLDSSPEPKILEVTYCP 840

RESULT 10
S48261
hypotheical protein YBR094w - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein YBR094w
C;Species: Saccharomyces cerevisiae
C;Date: 01-Aug-1995 #sequence_revision 11-Aug-1995 #text_change 09-Jul-2004
C;Accession: S48261; S45962; S41800; S44676
R;Mannhaupt, G.; Stucka, R.; Ehnl, S.; Vetter, I.; Feldmann, H.
Yeast 10, 1363-1381, 1994
A;Title: Analysis of a 70 kb region on the right arm of yeast chromosome II.
A;Reference number: S48255; MUID:95208357; PMID:7900426
A;Accession: S48261
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-753 <MAN>
A;Cross-references: UNIPROT:P38254; UNIPARC:UPI000013A3DD; EMBL:X78993; NID:G476045; PI
R;Feldmann, H.; Mannhaupt, G.; Schwarzlose, C.; Vetter, I.
submitted to the Protein Sequence Database, August 1994
A;Reference number: S45927
A;Accession: S45962
A;Molecule type: DNA
A;Residues: 1-753 <PE2>
A;Cross-references: UNIPARC:UPI000013A3DD; EMBL:Z35963; NID:G536366; PIDN:CAA85047.1; PID
R;Dekker, P.J.T.; Hoekert, W.; van Oosterum, K.; Grivell, L.A.
submitted to the EMBL Data Library, December 1992
A;Reference number: S31313
A;Accession: S41800
A;Molecule type: DNA
A;Residues: 167-351, 'TPE', 356-449, 'A', 451-562, 'R', 564-753 <DEK>
A;Cross-references: UNIPARC:UPI0000168D37; EMBL:X69881; NID:G4086; PIDN:CAA49508.1; PID
C;Genetics:
A;Cross-references: SGD:S00000298
A;Map position: 2R

Query Match 5.1%; Score 144.5; DB 2; Length 753;
Best Local Similarity 21.8%; Pred. No. 0.012;
Matches 57; Conservative 44; Mismatches 105; Indels 56; Gaps 9;

QY 140 WIIPKAASRGDRIVCMRVVEI-----LELAAADHP----- 171
DB 476 WIVKPSMSDKGGIRVFKEKTIEDIQAIFDSFDDESEAESGNDDDADDVNGEFMDNNKNV 535

QY 172 LSRDNKWWQKYIETPLLIC---DTKFDIROWFLVTDWNPDLTIWFYKESYLRFSTQRF-- 226
DB 536 ISQLRHFIIOEYLTNPILLASMDNRKFRICY--VVCGRDLQVFFVDRMLALFAAKFPVP 593

QY 227 -----SLDKLDSAIHLCNNAVQKYLKNDVGRSPILLPAHNMWTSRFO--EYLQROGRG 277
DB 594 LDPIYAVSVTDKLECHLTNTCLOQS-KKKDKDSVL-----EFDSEIEELPNERKS 642

QY 278 AVMGSVIYPSMKKAIHAAMKVAQDHVEPRKNSFELYGADFVLGRDPRPMLIENSSPTMH 337
DB 643 NI-KEQIHSITNDVFLAAVNVNRLNFOPLNFAETYGVDFLDSNVEVKLEINAFDPDK 701

QY 338 PSTPVTVAQLCAQVOEDTIKVAV 359
DB 702 QTGKDLKNLYDELFDFTVKYCV 723

A;Authors: Sqaers, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230

A;Accession: A70507
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-882 <COL>
A;Cross-references: UNIPROT:O33268; UNIPARC:UPI0000004F63; GB:Z97991; GB:AL123456; NID:9
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv0338c
C;Keywords: oxidoreductase

Query Match 4.2%; Score 120; DB 2; Length 882;
Best Local Similarity 27.2%; Pred. No. 0.98;
Matches 43; Conservative 11; Mismatches 52; Indels 52; Gaps 5;

QY 416 AQLKARGPSAMPD-----PAQGPPSPALQDLGLKEEKGPLALLAPLRGAES 465
DB 736 AKQQRAPAKAAPAAAPVTPVPAEAPQAPA-----PAAPAPVKGIGWA 782
QY 466 GGAAQPTRTKAAGKVELPACPCRHVS-----QAPNTGVPVPAQPAKSW 508
DB 783 AGAKRFGAKKAAPTPAAPAAPAVKGLGTAAGAKRPGAKKTPPPAPGLAEPAQAQPOEA 842
QY 509 DPNOLNAHLEP-----VLRLKTAEGALRPP 535
DB 843 KPQEPAAAPKPKQTDGDPAAAPAVKGLGIARGA-RPP 879

RESULT 14
S55659
segment protein 64 - equine herpesvirus 2
C;Species: equine herpesvirus 2
C;Date: 27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 09-Jul-2004
C;Accession: S55659
R;Reiford, E.A.R.; Watson, M.S.; Aird, H.C.; Perry, J.; Davison, A.J.
J. Mol. Biol. 249, 520-528, 1995
A;Title: The DNA sequence of equine herpesvirus 2.
A;Reference number: S55594; MUID:95302501; PMID:7783207
A;Accession: S55659
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-3436 <NEL>
A;Cross-references: UNIPROT:Q66666; UNIPARC:UPI0000007BF4; GB:U20824; NID:96951172; PIDN:
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1995

Query Match 4.2%; Score 119.5; DB 2; Length 3436;
Best Local Similarity 18.6%; Pred. No. 7.8; Mismatches 261; Indels 233; Gaps 29;
Matches 134; Conservative 94

QY 9 VVSHQSCSRSSRPRDQREAGS-----SDLSSRQDAENAEAKRLGQLQGLVDIAKVC 63
DB 1919 VAATGSLPRLRHSFAPCAEDTGAALFALSEAVKLEKSGRIRWRGVEGKLSDAHSLVA 1978
QY 64 QAYLGQLEHEDI-----DTSADAVEDLTEAEWEDLTQQYSLVHGDFAFTSNRYN 114
DB 1979 RSAAKGAQKRLFALIOGLRKDYAAQREIME-DWK-----AFVT----- 2018
QY 115 SQCOALLNRITSVNPQTDIDGLRNIIWIKPAKSRGDIVCMRVEIELELAADHPLSR 174
DB 2019 ---EAPINMEDVND-----ILRAAPNEAEFEAKKLEENREAREAKLEK 2064
QY 175 DNKVVVOKYIETPLLIC-----DTKFDIRQWFLVT--DWNPLTIWIFYKESYLFSTQRFSL 228
DB 2065 EAEEMLTQAVKRGQLQGWRIQNAFDNNAFGIITGEDWAAVAAEFQREGSLTSLTPGQL 2124
QY 229 DKLSAITH-----LCNNAVOKYLNVDVGRSP-----LLP--AH-NMTSTRFQBYLORQG 275
DB 2125 SKLTDKVEAEAEALLNKVSMPLNGPAFKPAPFDWLTTPVRAHNFYKLSFPLKLNROA 2184
QY 276 RG-----AVWGSVIYPS-----MKKAIAHAMKV---AQDHVPRKNSPELYGADF 317

DB 2185 EAVEAKMSQIEQAGADVYEAAGTGTLEAPVARALLRLRAARDEAAGLKGQIDEGRAY 2244
QY 318 VLG-----RDFRPWLIE----- 329
DB 2245 VQGVRAAGGEGPPAKPKAEIPKLLTYEQTSLANLPEDFQKNVLQNETLMNLQUREYL 2304
QY 330 -----INSSPTWHPST--PVTAAQLCAQVOEDTIKVAV-----DRSCDIGNFELLWR- 373
DB 2305 GRVTENINLSERAKTSRGEANARLAATVEENLPQANVISISSRLDKSDPVGFLGIVRD 2364
QY 374 QPVVEPPFSGSDCLCVAGSVRRARQVLPVC--NLKASASLLDQAFLKARGSAMPDPA 431
DB 2365 KOIVESDPYSATRESL--VMLHRTFKALLPLCPASLKRRLMELIGEBILREKG-----RG 2416
QY 432 QGPPSPALQ-----RDGLKE- 447
DB 2417 RGAPRYGVRAHETDDVGVLTAEIETLEARRVTGGKAAVEGWVKRDAYRNWVEDLALRSE 2476
QY 448 -EKGL-PL-----ALLAPLRGAESGGAQAQPTRTKAAGKVELPACPCRH----- 489
DB 2477 VEKRLGFLVEKSGREADADPLAVLQEEAAALLLEAKTGGLDK--SAPETHERVLELQMYL 2534
QY 490 -----VDSQAPNTGVPVPAQPAKSWDPNQLNAHLEPVLRLGLKTAEGALRPPPGGK 539
DB 2535 RFKLDPLKHYLDSQRVPFEAAPLSRALYMSNSQGERESRENSGEGEGVEGERRGEGGE 2594
QY 540 GS 541
DB 2595 GS 2596

RESULT 15
T41104
probable transcription factor subunit, TPR domain - fission yeast (Schizosaccharomyces
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C;Accession: T41104
R;Purnelle, B.; Goffeau, A.; Wood, V.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, September 1998
A;Reference number: Z21964
A;Accession: T41104
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-1006 <PUR>
A;Cross-references: UNIPROT:O74458; UNIPARC:UPI000006B686; EMBL:AL031535; PIDN:CAA20753
A;Experimental source: strain 972h-; cosmid c16C4
C;Genetics:
A;Gene: SPDB:SPCC16C4.14c
A;Map position: 3

Query Match 4.0%; Score 115.5; DB 2; Length 1006;
Best Local Similarity 21.0%; Pred. No. 2.6;
Matches 92; Conservative 58; Mismatches 137; Indels 151; Gaps 24;

QY 28 EEAGSDLSRRQD-----AENAEAKRLGLPG--QLVDIACKVCOA---VLQGLEHEDI 75
DB 3 QNGNSIVDSNMNETQNDTDFDAEQDNLNGYISEIVDEARNVSEVDKFLG----- 55
QY 76 DTSADAVEDLTEAEWEDLTQQYSLVHGDFAFTSNRYNFSQCOALLNRITSVNPQTDI-- 133
DB 56 DTSALQAEGL---WSDEESDYEG-----SDDESNSFSK-----TASRTEDDIAN 95
QY 134 -----DGLRNIIWIKPAKSRGR-----DIVCMRVEIELELAADHPLSRDNKW- 178
DB 96 EEWENLKAAGVFRK---VRKGHKGGRVSRADMLPSVEVQOQMLSL--AHNLFAQEGNFD 150
QY 179 VVQYIETPLLICDTKFDIRQWFLVTDNPLTIW-FYKESYLFSTQRFSLDKL-----DS 233
DB 151 EAQKLAEEIVRI-----DNNVIAAKWMLGECHQRGNGRVNIEKCLIAWMA 196
QY 234 AIHLCNNAVOKYLNVDVGRSPLPAHNMW-TSTRFQBYLORQGRGAVWGSVIY----- 285
DB 197 AAHL-----KP--KDHELWFTCAKLSSELE-----FWDQADYCNRAVS 233

```

QY 286 -----PSMKKATAHAKVAQDHVEPRKNSFELYGADFVLGRDRFREWLIENSPTMHPS 339
Db 234 AKPPKSELKKYIWNRSVLNKEHGLKAA---EGFKLL-----QS 272
QY 340 TPVTAOLCAQVOEDTIKVAVDRSCDIGNFELLWRQPVVEPPFSGSDLCVAGSVRRARR 399
Db 273 SPYNASILKNLAEIYIKIHAPREI-LKQFEIAWKYFYQYPAPPIGNDIF----- 320
QY 400 QVLPVCNLRKASLDDAQ 417
Db 321 -DLPTLNLVLAELLDDHQ 337

```

Search completed: April 4, 2006, 12:51:51
Job time : 24.1115 secs

THIS PAGE BLANK (USPTO)

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	2854	100.0	541	8	ADJ933358	Human BGS
2	2843	99.6	541	8	ADJ933365	Human tubu
3	1560	54.7	293	8	ADJ933366	Human BGS
4	1299	45.5	242	8	ADJ933360	Human BGS
5	1043.5	36.6	744	4	AA894796	Human pro
6	1013	35.5	488	5	AAU74334	Human cyt
7	995.5	34.9	399	6	ABU11512	Human MDD
8	969.5	34.0	362	3	AA858909	Breast ann
9	956.5	33.5	352	3	AA943005	Human ORP
10	956.5	33.5	352	4	AA839450	Human pol
11	956.5	33.5	352	8	ADJ933457	Human HOT
12	956.5	33.5	352	8	ABM80420	Human HOT
13	944.5	33.1	326	7	ADJ93455	Tumour-as
14	865.5	30.3	292	8	ADJ93455	Human pro
15	722.5	25.3	992	4	AB865645	Human HOT
16	715	25.1	432	6	AB865645	Drosophil
17	536.5	18.8	719	4	AB920448	Human pro
18	524.5	18.4	281	8	AB865541	Drosophil
19	519	18.2	362	4	AQ666114	Novel hum
20	516	18.1	402	8	AAW41236	Human pol
21	375.5	13.2	566	4	ADJ02747	Novel hum
22	336	11.8	160	8	ABG05971	Novel hum
23	330	11.6	347	6	ADH45424	Human mol
24	328	11.5	496	4	ABU00150	Human nov
					AB64074	Drosophil

PS	Claim 5; SEQ ID NO 2; 343pp; English.	
XX	This invention relates to a novel testis-specific tubulin tyrosine-ligase	
CC	-like polypeptide, designated the BGS-42 polypeptide. The invention may	
CC	be useful for the development of compounds with a cytosstatic, respiratory	
CC	-Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,	
CC	antiinflammatory, anabolic, hypertensive, osteopathic, nootropic,	
CC	antiparkinsonian, antiarthritic, antiaesthatic, anti-HIV, antibacterial,	
CC	immunosuppressive, antiseborrheic or dermatological activity acting as	
CC	tyrosine ligase modulators. In addition, the disclosed sequences may be	
CC	useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be	
CC	used for diagnosing a pathological condition or a susceptibility to a	
CC	pathological condition in a subject, and for preventing, treating or	
CC	ameliorating a medical condition, such as a disorder related to aberrant	
CC	tubulin ligase activity, a disorder related to aberrant tubulin-	
CC	carboxypeptidase activity, aberrant cellular proliferation, reproductive	
CC	disorders, testicular disorders, testicular cancer, pulmonary disorders,	
CC	lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,	
CC	neural disorders, brain cancer, liver cancer, or proliferative condition	
CC	of the testis, lung, small intestine, brain or lymph tissue. The BGS-42	
CC	polypeptide, polynucleotide, or their modulators are also useful for	
CC	treating infertility, Cushing's syndrome, emphysema, Addison's disease,	
CC	disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS	
CC	-42 polypeptide can be used as a preventive agent for immunological	
CC	disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's	
CC	disease or scleroderma. The antibodies may be used to purify, detect and	
CC	target the BGS-42 polypeptides. The present sequence is that of the human	
CC	BGS-42 protein of the invention.	
XX	Sequence 541 AA;	
SQ	Query Match	
	Best Local Similarity 100.0%; Score 2854; DB 8; Length 541;	
	Matches 541; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1 MASSILKVVVSHQSCSSRSKRDQREAGSDLSRQDAENAEAKLRLPGQLVDIAC 60	
DB	1 MASSILKVVVSHQSCSSRSKRDQREAGSDLSRQDAENAEAKLRLPGQLVDIAC 60	
QY	61 KVCQAYLGQLEHEDIDTSADAVEDLTEAEWEDLTQQYSLVHGDAFISNRSNYFSQCOAL 120	
DB	61 KVCQAYLGQLEHEDIDTSADAVEDLTEAEWEDLTQQYSLVHGDAFISNRSNYFSQCOAL 120	
QY	121 LNRITSNVPTDIDGLRNIMIIPAKSKRGDIIVCMRVREIELELAADHPLSRDNKVV 180	
DB	121 LNRITSNVPTDIDGLRNIMIIPAKSKRGDIIVCMRVREIELELAADHPLSRDNKVV 180	
QY	181 QKIETPLLCDTKFDIRQWFLVTDWNPFLTIFWKESYLVRFSTORFSLDKLDSAIHLCNN 240	
DB	181 QKIETPLLCDTKFDIRQWFLVTDWNPFLTIFWKESYLVRFSTORFSLDKLDSAIHLCNN 240	
QY	241 AVQYKLVNDVGRSPLPAHNWMTSTRFQYLRQGRGAVMGSVITYPSMKKAIAHAMKVAQ 300	
DB	241 AVQYKLVNDVGRSPLPAHNWMTSTRFQYLRQGRGAVMGSVITYPSMKKAIAHAMKVAQ 300	
QY	301 DHVEPRKNSFELYGADFVLGRDPRPWLIEINSSPTMHPSTPTAQLCAQVOEDTIKVAVD 360	
DB	301 DHVEPRKNSFELYGADFVLGRDPRPWLIEINSSPTMHPSTPTAQLCAQVOEDTIKVAVD 360	
QY	361 RSCDIGNFELLWRQPVVPEPPFSGSLCVAGSVRRARROVLPVCNLKASASLLDAOPLK 420	
DB	361 RSCDIGNFELLWRQPVVPEPPFSGSLCVAGSVRRARROVLPVCNLKASASLLDAOPLK 420	
QY	421 ARGPSAMPDPAQGPSPALQRDGLGKEEGLPLALLAPLRGAESGGAQAQTRTKAAGV 480	
DB	421 ARGPSAMPDPAQGPSPALQRDGLGKEEGLPLALLAPLRGAESGGAQAQTRTKAAGV 480	
QY	481 ELPACPCRHVDSQAPNTGVPVQAQPAKSWDPNQLNAHPLIEPVLRGLKTAEGALRPPPGKG 540	
DB	481 ELPACPCRHVDSQAPNTGVPVQAQPAKSWDPNQLNAHPLIEPVLRGLKTAEGALRPPPGKG 540	
QY	541 \$ 541	

DB	541 S 541	
RESULT 2		
ADJ93365		
ID	ADJ93365 standard; protein; 541 AA.	
XX		
AC	ADJ93365;	
XX		
DT	06-MAY-2004 (first entry)	
XX		
DE	Human tubulin tyrosine ligase protein consensus sequence SeqID13.	
XX		
KW	testis-specific tubulin tyrosine-ligase-like polypeptide;	
KW	BGS-42 polypeptide; cytosstatic; respiratory-Gen; gastrointestinal-Gen;	
KW	neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;	
KW	osteopathic; nootropic; antiparkinsonian; antiarthritic; antiaesthatic;	
KW	anti-HIV; antibacterial; immunosuppressive; antiseborrheic;	
KW	dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;	
KW	tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;	
KW	testicular disorder; testicular cancer; pulmonary disorder; lung cancer;	
KW	gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;	
KW	brain cancer; liver cancer; proliferative condition; testis; lung;	
KW	small intestine; brain; lymph tissue; infertility; Cushing's syndrome;	
KW	emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;	
KW	Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;	
KW	sepsis; acne; Sjogren's disease; scleroderma; human.	
OS	Homo sapiens.	
OS	Synthetic.	
XX		
PN	WO2004005487-A2.	
XX		
PD	15-JAN-2004.	
XX		
PF	09-JUL-2003; 2003WO-US021605.	
XX		
PR	09-JUL-2002; 2002US-0394725P.	
XX		
PA	(BRIM) BRISTOL-MYERS SQUIBB CO.	
XX		
PI	Feder JN, Wu S, Nelson TC;	
XX		
DR	WFI; 2004-099381/10.	
DR	N-PSDB; ADJ93364.	
XX		
PT	New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,	
PT	useful for preventing, treating or ameliorating a medical condition, e.g.	
PT	aberrant cellular proliferation, reproductive disorders or testicular	
PT	disorders.	
XX		
PS	Example 4; SEQ ID NO 13; 343pp; English.	
XX		
CC	This invention relates to a novel testis-specific tubulin tyrosine-ligase	
CC	-like polypeptide, designated the BGS-42 polypeptide. The invention may	
CC	be useful for the development of compounds with a cytosstatic, respiratory	
CC	-Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,	
CC	antiinflammatory, anabolic, hypertensive, osteopathic, nootropic,	
CC	antiparkinsonian, antiarthritic, antiaesthatic, anti-HIV, antibacterial,	
CC	immunosuppressive, antiseborrheic or dermatological activity acting as	
CC	tyrosine ligase modulators. In addition, the disclosed sequences may be	
CC	useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be	
CC	used for diagnosing a pathological condition or a susceptibility to a	
CC	pathological condition in a subject, and for preventing, treating or	
CC	ameliorating a medical condition, such as a disorder related to aberrant	
CC	tubulin ligase activity, a disorder related to aberrant tubulin-	
CC	carboxypeptidase activity, aberrant cellular proliferation, reproductive	
CC	disorders, testicular disorders, testicular cancer, pulmonary disorders,	
CC	lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,	
CC	neural disorders, brain cancer, liver cancer, or proliferative condition	
CC	of the testis, lung, small intestine, brain or lymph tissue. The BGS-42	
CC	polypeptide, polynucleotide, or their modulators are also useful for	
CC	treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's	

CC disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
CC -42 polypeptide can be used as a preventive agent for immunological
CC disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
CC disease or scleroderma. The antibodies may be used to purify, detect and
CC target the BGS-42 polypeptides. The present sequence is that of the
CC tubulin tyrosine ligase protein consensus sequence which was used in the
CC exemplification of the invention.

XX
SQ Sequence 541 AA;

Query Match 99.6%; Score 2843; DB 8; Length 541;
Best Local Similarity 99.6%; Pred. No. 4.8e-263;
Matches 539; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MASSILKVVVSHQSCSRSSRPRDREAGSDLSRRDAENAEAKURLPGQLVDIAC 60
Db 1 MASSILKVVVSHQSCSRSSRPRDREAGSDLSRRDAENAEAKURLPGQLVDIAC 60
QY 61 KVCQAVLGQLEHEDITSADAVEDLTEAWEDLTQYYSLVHGDAFINSRNYFSQCQAL 120
Db 61 KVCQAVLGQLEHEDITSADAVEDLTEAWEDLTQYYSLVHGDAFINSRNYFSQCQAL 120
QY 121 LNRITSVNPQTDIDGLRNIWIIPAKSRGRDIVCMRVVEIELEAAADHPLSRDNKVV 180
Db 121 LNRITSVNPQTDIDGLRNIWIIPAKSRGRDIVCMRVVEIELEAAADHPLSRDNKVV 180
QY 181 QKIETPLLICDTKFDIROWFLVTDWNPLTIWFKESYLRFSTQRFSLDKLDSAIHLN 240
Db 181 QKIETPLLICDTKFDIROWFLVTDWNPLTIWFKESYLRFSTQRFSLDKLDSAIHLN 240
QY 241 AVQKYLKNDVGRSPLPAHNMWTSRFOYLQROGRGAVGWSVIYPSMKKAIAHAMKVAQ 300
Db 241 AVQKYLKNDVGRSPLPAHNMWTSRFOYLQROGRGAVGWSVIYPSMKKAIAHAMKVAQ 300
QY 301 DHVEPRKNSFELYGADFVLGRDPRWLTIEINSSPTWHPSTPTAQLCAQVQEDTTIKVAVD 360
Db 301 DHVEPRKNSFELYGADFVLGRDPRWLTIEINSSPTWHPSTPTAQLCAQVQEDTTIKVAVD 360
QY 361 RSCDIGNFELLWRQVPVPPPPSGSDLCVAGSVRRARRQVLPVCLNKASALLDAQPLK 420
Db 361 RSCDIGNFELLWRQVPVPPPPSGSDLCVAGSVRRARRQVLPVCLNKASALLDAQPLK 420
QY 421 ARGPSAMPDPAQGPSALORDLGLKEKGLPLALLAPLARGAESGGAQPTRTVAAGKV 480
Db 421 ARGPSAMPDPAQGPSALORDLGLKEKGLPLALLAPLARGAESGGAQPTRTVAAGKV 480
QY 481 ELPACPCRHRVDSQAPNTGVVPAQPAKSDPNQNAHPLPVLRLKTAEGALRPPPGGKG 540
Db 481 ELPACPCRHRVDSQAPNTGVVPAQPAKSDPNQNAHPLPVLRLKTAEGALRPPPGGKG 540
QY 541 S 541
Db 541 S 541

RESULT 3
ADJ93366 standard; protein; 293 AA.

XX AC ADJ93366;

XX DT 06-MAY-2004 (first entry)

XX DE Human BGS-42 protein-related TTL1 domain.

XX testis-specific tubulin tyrosine-ligase-like polypeptide;
KW BGS-42 polypeptide; cytosolic; respiratory-gen; gastrointestinal-gen;
KW neuroprotective; endocrine-gen; antiinflammatory; anabolic; hypertensive;
KW osteopathic; nontropic; antiparkinsonian; antiarthritic; antiasthmatic;
KW anti-HIV; antibacterial; immunosuppressive; antiseborrheic;
KW dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;
KW tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;
KW testicular disorder; testicular cancer; pulmonary disorder; lung cancer;

KW gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;
KW brain cancer; liver cancer; proliferative condition; testis; lung;
KW small intestine; brain; lymph tissue; infertility; Cushing's syndrome;
KW emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;
KW Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;
KW sepsis; acne; Sjogren's disease; scleroderma; human; TTL1 domain.

OS Homo sapiens.

XX WO2004005487-A2.

XX 15-JAN-2004.

PF 09-JUL-2003; 2003WO-US021605.

PR 09-JUL-2002; 2002US-0394725P.

XX (BRIM) BRISTOL-MYERS SQUIBB CO.

XX Feder JN, Wu S, Nelson TC;

XX WPI; 2004-099381/10.

XX New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,
XX useful for preventing, treating or ameliorating a medical condition, e.g.
XX aberrant cellular proliferation, reproductive disorders or testicular
XX disorders.

PS Disclosure; SEQ ID NO 14; 343pp; English.

XX This invention relates to a novel testis-specific tubulin tyrosine-ligase
XX -like polypeptide, designated the BGS-42 polypeptide. The invention may
XX be useful for the development of compounds with a cytostatic, respiratory
XX -Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,
XX antiinflammatory, anabolic, hypertensive, osteopathic, nontropic,
XX antiparkinsonian, antiarthritic, antiasthmatic, anti-HIV, antibacterial,
XX immunosuppressive, antiseborrheic or dermatological activity acting as
XX tyrosine ligase modulators. In addition, the disclosed sequences may be
XX useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be
XX used for diagnosing a pathological condition or a susceptibility to a
XX pathological condition in a subject, and for preventing, treating or
XX ameliorating a medical condition, such as a disorder related to aberrant
XX tubulin ligase activity, a disorder related to aberrant tubulin-
XX carboxypeptidase activity, aberrant cellular proliferation, reproductive
XX disorders, testicular disorders, testicular cancer, pulmonary disorders,
XX lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
XX neural disorders, brain cancer, liver cancer, or proliferative condition
XX of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
XX polypeptide, polynucleotide, or their modulators are also useful for
XX treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
XX disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
XX -42 polypeptide can be used as a preventive agent for immunological
XX disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
XX disease or scleroderma. The antibodies may be used to purify, detect and
XX target the BGS-42 polypeptides. The present sequence is that of the TTL1
XX domain of the human BGS-42 protein of the invention.

XX SQ Sequence 293 AA;

Query Match 54.7%; Score 1560; DB 8; Length 293;

Best Local Similarity 100.0%; Pred. No. 2.2e-140;

Matches 293; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 73 EDITSADAVEDLTEAWEDLTQYYSLVHGDAFINSRNYFSQCQALLNRTSVNPQTD 132

Db 1 EDITSADAVEDLTEAWEDLTQYYSLVHGDAFINSRNYFSQCQALLNRTSVNPQTD 60

QY 133 IDGLRNIWIIPAKSRGRDIVCMRVVEIELEAAADHPLSRDNKVVQKIETPLLICD 192

Db 61 IDGLRNIWIIPAKSRGRDIVCMRVVEIELEAAADHPLSRDNKVVQKIETPLLICD 120

QY 193 TKFDIROWFLVTDWNPLTIWFKESYLRFSTQRFSLDKLDSAIHLNNAVQKYLKNDVGR 252

Db 121 TKFDIRQWFLVTDNPLTIWFKESYLRFSTQFSLDKLDSAHLCCNNAVQKYLKNDVGR 180
Qy 253 SPLLPAHNNMTSTRFOEYLQROGRGAVGSGVIYPSMKKAIAHAMKVAQDHVEPRKNSFEL 312
Db 181 SPLLPAHNNMTSTRFOEYLQROGRGAVGSGVIYPSMKKAIAHAMKVAQDHVEPRKNSFEL 240
Qy 313 YGADFVLGRDPRFLWLEINSSPTMHPSTPVTALCAQVQEDTIKVAVDKSCDI 365
Db 241 YGADFVLGRDPRFLWLEINSSPTMHPSTPVTALCAQVQEDTIKVAVDKSCDI 293

RESULT 4
ADJ93360
ID ADJ93360 standard; protein; 242 AA.
AC ADJ93360;
DT 06-MAY-2004 (first entry)
XX
XX Human BGS-42 protein sequence SeqID4.
DE
XX testis-specific tubulin tyrosine-ligase-like polypeptide;
KW BGS-42 polypeptide; cytosolic; respiratory-Gen; gastrointestinal-Gen;
KW neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;
KW osteopathic; nontropic; antiparkinsonian; antiarthritic; antiasthmatic;
KW anti-HIV; antibacterial; immunosuppressive; antiseborrheic;
KW dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;
KW tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;
KW testicular disorder; testicular cancer; pulmonary disorder; lung cancer;
KW gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;
KW brain cancer; liver cancer; proliferative condition; testis; lung;
KW small intestine; brain; lymph tissue; infertility; Cushing's syndrome;
KW emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;
KW Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;
KW sepsis; acne; Sjogren's disease; scleroderma; human.
XX
XX Homo sapiens.
XX WO2004005487-A2.
XX 15-JAN-2004.
XX 09-JUL-2003; 2003WO-US021605.
XX 09-JUL-2002; 2002US-0394725P.
XX (BRIM) BRISTOL-MYERS SQUIBB CO.
XX Feder JN, Wu S, Nelson TC;
XX WPI; 2004-099381/10.
XX
XX New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,
XX useful for preventing, treating or ameliorating a medical condition, e.g.
XX aberrant cellular proliferation, reproductive disorders or testicular
XX disorders.
XX Disclosure; SEQ ID NO 4; 343pp; English.
XX
XX This invention relates to a novel testis-specific tubulin tyrosine-ligase
XX -like polypeptide, designated the BGS-42 polypeptide. The invention may
XX be useful for the development of compounds with a cytostatic, respiratory
XX -Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,
XX antiinflammatory, anabolic, hypertensive, osteopathic, nontropic,
XX antiparkinsonian, antiarthritic, antiasthmatic, anti-HIV, antibacterial,
XX immunosuppressive, antiseborrheic or dermatological activity acting as
XX tyrosine ligase modulators. In addition, the disclosed sequences may be
XX useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be
XX used for diagnosing a pathological condition or a susceptibility to a
XX pathological condition in a subject, and for preventing, treating or
XX ameliorating a medical condition, such as a disorder related to aberrant
XX tubulin ligase activity, a disorder related to aberrant tubulin-
XX carboxypeptidase activity, aberrant cellular proliferation, reproductive

CC disorders, testicular disorders, testicular cancer, pulmonary disorders,
CC lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
CC neural disorders, brain cancer, liver cancer, or proliferative condition
CC of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
CC polypeptide, polynucleotide, or their modulators are also useful for
CC treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
CC disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
CC -42 polypeptide can be used as a preventive agent for immunological
CC disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
CC disease or scleroderma. The antibodies may be used to purify, detect and
CC target the BGS-42 polypeptides. The present sequence is that of the human
CC BGS-42 protein (partial sequence) of the invention.
XX
XX Sequence 242 AA;
Qy 133 IDGLRNIIWKPAKSRGRDIVCMRVEILELAADHPLSRDNKVVQKIETPLLICD 192
Db 1 IDGLRNIIWKPAKSRGRDIVCMRVEILELAADHPLSRDNKVVQKIETPLLICD 60
Qy 193 TKFDIRQWFLVTDNPLTIWFKESYLRFSTQFSLDKLDSAHLCCNNAVQKYLKNDVGR 252
Db 61 TKFDIRQWFLVTDNPLTIWFKESYLRFSTQFSLDKLDSAHLCCNNAVQKYLKNDVGR 120
Qy 253 SPLLPAHNNMTSTRFOEYLQROGRGAVGSGVIYPSMKKAIAHAMKVAQDHVEPRKNSFEL 312
Db 121 SPLLPAHNNMTSTRFOEYLQROGRGAVGSGVIYPSMKKAIAHAMKVAQDHVEPRKNSFEL 180
Qy 313 YGADFVLGRDPRFLWLEINSSPTMHPSTPVTALCAQVQEDTIKVAVDKSCDI 372
Db 181 YGADFVLGRDPRFLWLEINSSPTMHPSTPVTALCAQVQEDTIKVAVDKSCDI 240
Qy 373 RQ 374
Db 241 RQ 242

RESULT 5
AAB94796
ID AAB94796 standard; protein; 744 AA.
XX
XX AAB94796;
XX
XX 26-JUN-2001 (first entry)
XX Human protein sequence SEQ ID NO:15921.
XX
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX Homo sapiens.
XX
XX EP1074617-A2.
XX
XX 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-00116126.
XX
XX 29-JUL-1999; 99JP-00248036.
XX 27-AUG-1999; 99JP-0030253.
XX 11-JAN-2000; 2000JP-00118776.
XX 02-MAY-2000; 2000JP-00183767.
XX 09-JUN-2000; 2000JP-00241899.
XX (HELI-) HELIX RES INST.
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
XX

PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.

XX
PS Claim 8; SEQ ID NO 15921; 2537pp + Sequence Listing; English.

XX The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dr primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention

XX Sequence 744 AA;

Query Match 36.6%; Score 1043.5; DB 4; Length 744;
Best Local Similarity 40.5%; Pred. No. 3.2e-90;
Matches 234; Conservative 85; Mismatches 192; Indels 67; Gaps 12;

QY 2 ASSILKVVVSHQSCRSRKRDRQREAGSDLSRQDAENAEAKLGLPQLVDIACK 61
DB 151 ARNLVLKVVKSE-----WKSYIQIAVEEASGDKQPKKNPVL-----VSPFEFDEALC 201
QY 62 VQOAVLGOLEHEDIDTSADAVEDLTAEWEDLTQYVSLVHGDAFISNSRNFYSCQALL 121
DB 202 ACEEVLNLAHWDIDKLEAPLYLTPGWSLFLQYVQVHGEALRHLDLTQVQCEDIL 261
QY 122 NRITSVNPQTDIDGLRNWIIPKAASRGDRIVCMRDVEEILELAADHPLSRDNKVVQ 181
DB 262 QQLQAVVQIDMEGRNIWIVKPGAKSRGRGIMCMWHLLEMLKLVNGNPVVMKQGVVQ 321
QY 182 KYIETPLICDTKFDIRQWFLVTDNPLTIYKESYLRFTQRFSLQKLSAHLICNNA 241
DB 322 KYIERPLLI FGTFKFLQWFLVTDNPLTVWFYRDSYIRFTQPFSLKNDLSVHLNNS 381
QY 242 VQYLLKNDVGRSPLLAHNMWITSTRFOELQQRGAVMGSVIYPSMKKATAHAMKVAQD 301
DB 382 IQKLENSCHRRHPLDPDNMSSQRFQAHQWEGMAPNASTIIVPGMDAVIHALQTSQD 441
QY 302 HVEPRKNSFELYGADFLVGRDPRPWLIEINSPTMHPSTPTVAQLCAOVQEDTIKVAV-- 359
DB 442 TVQCRKASFELYGADFVGFQFQWPLIEINASTPTWAPSTAVTARLCAGVQADTLRVVIDR 501
QY 360 --DRSCDIGNFELLWRQVPPPPPSGSDLCVAGVSVRRARRQVLPCVNLKASALLDAQ 417
DB 502 MLDNRCDTGAFELIYKPAVEVPQVVGIRLLVEGFTIKK-----PMANCHRRMGRVDAV 555
QY 418 PLKARGPS-----AMPDPAQGPSPALORDLGL-----KEEKG 450
DB 556 PLLTORGGEARHHPSSLHTKAO--LPSPHVLRHQGVLRROHSLKLVGFKALSTTGKALRT 614
QY 451 LPLA----LLAPLRGAESGGAOPTRTKA-----AGKVELPACPCRHVDSQ--AP-NTG 498
DB 615 LPTAKVFISLPNLDKFAVPSILKPKAPALLCLRGPOLEVPCCCLPKSEQFLAPVGRS 674
QY 499 VFVAQPAKSWDPNQLNAPHLFVLRGLKTAEGALRPPP 536

DB 675 RPKANRRPCDCKPRAEACPMKRL-----SPLKPLP 704

RESULT 6
AAU74334

ID AAU74334 standard; protein; 488 AA.

XX AAU74334;

DT 12-MAR-2002 (first entry)

DE Human cytoskeleton-associated protein (CYSKP) #5.

KW Human; cytoskeleton-associated protein; CYSKP; autoimmune disorder;
cell proliferative disorder; inflammatory disorder; prion disease;
vesicle trafficking disorder; gastrointestinal disorder; muscle disorder;
neurological disorder; cell motility disorder; reproductive disorder;
spinal cord disease; central nervous system disorder; mental disorder;
gene therapy; cancer.

OS Homo sapiens.

XX WO200185942-A2.

XX 15-NOV-2001.

XX 03-MAY-2001; 2001WO-US014355.

XX 05-MAY-2000; 2000US-0201960P.

XX 08-MAY-2000; 2000US-0202729P.

XX 05-JUN-2000; 2000US-0209705P.

XX 07-JUN-2000; 2000US-0210149P.

XX 21-JUN-2000; 2000US-0213215P.

XX (INCY-) INCYTE GENOMICS INC.

XX Yue H, Tang YT, Au-Young J, Lu DAM, Baughn MR, Hillman JL;
PI Azimzai Y, Lal P, Yao MG, Bandman O, Burford N, Batra S, Kearney L;
PI Policky JL;

XX WPI; 2002-062248/08.

XX N-PSDB; AAS99894.

XX New cytoskeleton-associated proteins and polynucleotides, useful for
diagnosing, preventing and treating cell proliferative, autoimmune,
inflammatory, neurological, cell motility, reproductive and muscle
disorders.

XX Claim 1; Page 130-131; 194pp; English.

XX The invention relates to human cytoskeleton-associated polypeptides
(CYSKP) and their associated polynucleotide sequences. The sequences are
useful in the treatment of disorders associated with overexpression or
underexpression of CYSKP in a patient. The disorders include cell
proliferative disorders (such as cancer, actinic keratosis,
arteriosclerosis, cirrhosis, hepatitis and psoriasis),
autoimmune/inflammatory disorders (such as, asthma, atherosclerosis,
osteoporosis, Crohn's disease, rheumatoid arthritis, diabetes mellitus
and anaemia), vesicle trafficking disorders (such as
hypercholesterolaemia, diabetes insipidus, Grave's disease and goitre),
gastrointestinal disorders, prion diseases, neurological disorders (such
as epilepsy, stroke, cerebral neoplasms, Alzheimer's disease,
Huntington's disease, Parkinson's disease, amyotrophic lateral sclerosis
and other motor neuron disorders), cell motility disorders, reproductive
disorders (such as endometriosis and polycystic ovary syndrome), muscle
disorders (such as myocarditis, migraine, hypertension, hypoglycaemia,
myocardial infarction, epilepsy and muscular dystrophy), spinal cord
diseases, central nervous system disorders (such as Down syndrome and
cerebral palsy) and mental disorders (such as anxiety and schizophrenia).

XX Sequences AAU74330-AAU74363 represent human CYSKP of the invention
Sequence 488 AA;

```
Query Match      35.5%; Score 1013; DB 5; Length 488;
Best Local Similarity 45.5%; Pred. No. 1.4e-87;
Matches 214; Conservative 68; Mismatches 138; Indels 50; Gaps 10;

QY 74 DITSDADVEDLTEAEWEDLTQOYSLVHGDAFISNRVFSQOALLNRITSVNPQTDI 133
DB 2 DIDKDLEAPLYLPEGWSFLQRYOVVHGEALRHLDTVQVQCEDILOQLQAVVQIDM 61

QY 134 DGLNRNIWKPAKSRGRDVCMDRVEEILELAADHPLSRDKWVQVYIETPLFICDT 193
DB 62 EGRNIWIKPGAKSRGRGIMCNDHLEMLKLVNGPNVWKQGVVQVYIERPLIFGT 121

QY 194 KFDIRQWFLVDNPLTITWPKSYLRFSTQRFSLDKLSAHLCNNAVOKYLNKDVGRS 253
DB 122 KFDLRQWFLVDNPLTITWPKSYLRFSTQRFSLDKLSAHLCNNAVOKYLNKDVGRS 181

QY 254 PLLPAHNMWTSRFOYLOQGRGAVGWSVIFYPSMKKATAHAKVQADHVEPRKNSFELY 313
DB 182 PLLPPDNMWSQRFOAHLQEMGAPNAWSTIIVFGMKDAVIHALQTSQDTVQCRKASFELY 241

QY 314 GADFVLGRDPRPWLIEINSSPTMHPSPVTAQLCAVOQEDTIKAVV-----DRSCDIGNFE 369
DB 242 GADFVGEDFQFWLIEINASSPTMAPSTAVTARLCAGVQADTLAVVIDRDLNCDTGAPE 301

QY 370 LLWRQPVVEPPFPFGSDLCVAGSVRRARQVLPVNCNLKASLLDAQPLKARGPSAMPD 429
DB 302 LIYKQPAVEPVQVIGIRLLVEGTIKK-----PMAMCH 334

QY 430 PAQG--PPSPAL--ORDGLKEEGLPLALLAPLR--GAESGGAAQP---TRTKAAGK-- 479
DB 335 RRMGVRPAVPLLTQSGGEGKDSGIPHRSASRKGRTGARSGLHSEKPVSTATTSAPECKG 394

QY 480 ----VELPACFCRHHV-DSQAPNT--GVPVAQPAKSWD--PNQLNAHPLEP 520
DB 395 KGKAKRATALVCPNLWENDAPSTRMGCIFTWTFSSGDRQPHHLNRPLSP 444

RESULT 7
ABU11512
ID ABU11512 standard; protein; 399 AA.
XX
AC ABU11512;
XX
XX 12-FEB-2003 (first entry)
XX
XX Human MDTT polypeptide SEQ ID 459.
XX
XX MDTT; human; disease detection and treatment molecule polypeptide;
XX anti-inflammatory; immunosuppressive; osteopathic; cytostatic; anti-HIV;
XX haemostatic; nephrotropic; antianaemic; antipsoriatic; hepatotropic;
XX gene therapy; protein replacement therapy; cell proliferative disorder;
XX cancer; adenocarcinoma; leukaemia; lymphoma; melanoma; myeloma; sarcoma;
XX anaemia; Crohn's disease; acquired immunodeficiency syndrome; AIDS;
XX Goodpasture's syndrome; inflammation; osteoporosis; thrombocytopaenia;
XX psoriasis; hepatitis.
XX
OS Homo sapiens.
XX
XX WO200279449-A2.
XX
XX 10-OCT-2002.
XX
XX 27-MAR-2002; 2002WO-US009944.
XX
XX 28-MAR-2001; 2001US-0279619P.
XX 29-MAR-2001; 2001US-0280067P.
XX 29-MAR-2001; 2001US-0280068P.
XX 16-MAY-2001; 2001US-0291280P.
XX 17-MAY-2001; 2001US-0291829P.
XX 17-MAY-2001; 2001US-0291849P.
XX 19-JUN-2001; 2001US-0299428P.
XX 20-JUN-2001; 2001US-0299776P.

PR 20-JUN-2001; 2001US-0300001P.
XX
XX (INCY-) INCYTE GENOMICS INC.
XX
XX Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;
XX Dufour GE, Hillman JL, Yu JY, Tuason O, Yap PE, Amshay SR;
XX Daugherty SC, Dam TC, Liu TF, Nguyen DA, Kleefeld Y, Gerstin EH;
XX Peralta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B;
XX Flores V, Marwaha R, Lo A, Lan RY, Urashka ME;
XX
XX WPI; 2003-058431/05.
XX
XX N-PSDB; ABX34502.
XX
XX New purified disease detection and treatment molecule proteins and
XX polynucleotides, useful for diagnosing, treating or preventing cancers
XX (e.g. leukemia or sarcoma), anemia, Crohn's disease, AIDS, osteoporosis
XX or hepatitis.
XX
XX Claim 27; SEQ ID NO 459; 339pp + Sequence Listing; English.
XX
XX This invention describes a novel disease detection and treatment molecule
XX polypeptide (MDTT) which has anti-inflammatory, immunosuppressive,
XX osteopathic, cytostatic, anti-HIV, haemostatic, nephrotropic,
XX antianaemic, antipsoriatic and hepatotropic activity. The polynucleotides
XX and the polypeptides of the invention can be used for gene therapy,
XX protein replacement therapy and are useful for treating a variety of
XX diseases or conditions. These polypeptides or polynucleotides are
XX particularly useful for diagnosing, treating or preventing cell
XX proliferative disorders (e.g. cancers including adenocarcinoma,
XX leukaemia, lymphoma, melanoma, myeloma or sarcoma), anaemia, Crohn's
XX disease, acquired immunodeficiency syndrome (AIDS), Goodpasture's
XX syndromes, inflammation, osteoporosis, thrombocytopaenia, psoriasis or
XX hepatitis. ABU11450-ABU11845 represent the MDTT polynucleotides encoded
XX by ABU11450-ABU11845, described in the disclosure of the invention. NOTE:
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 399 AA;
XX
XX Query Match      34.9%; Score 995.5; DB 6; Length 399;
XX Best Local Similarity 53.4%; Pred. No. 4.9e-86;
XX Matches 187; Conservative 57; Mismatches 101; Indels 5; Gaps 2;

QY 54 QLVDIACKVQAVYLGQLEHEDITSADAVEDLTEAEWEDLTQOYSLVHGDAFISNRNY 113
DB 29 EFVDEALCACEEYLSNLAHMDIDKLEAPLYLTPGWSLFQRYQVYVHGEALRHLDTQ 88

QY 114 FSCQALLNRITSVNPOTDIDGLNRNIWKPAKSRGRDVCMDRVEEILELAADHPLS 173
DB 89 VQCEDILOQLQAVVQIDMEGRNIWIKPGAKSRGRGIMCNDHLEMLKLVNGPNVPM 148

QY 174 RDNKVVVQKYIETPLITCDTKFDIRQWFLVDNPLTITWPKSYLRFSTQRFSLDKLDS 233
DB 149 KDGKVVVQKYIETPLIFGTGKFDLRQWFLVDNPLTITWPKSYLRFSTQRFSLDKLDS 208

QY 234 AIHLCNNAVOKYLNKDVGRSPLIPAHNMWTSRFOYLOQGRGAVGWSVIFYPSMKKATA 293
DB 209 SVHLCNNSIQKHLNESCRRHPFLPPDNMWSQRFOAHLQEMGAPNAWSTIIVFGMKDAVI 268

QY 294 HAMKVAQDHVEPRKNSFELYGADFVLGRDPRPWLIEINSSPTMHPSPVTAQLCAVOQED 353
DB 269 HALQTSQDTVQCRKASFELYGADFVGEDFQFWLIEINASSPTMAPSTAVTARLCAGVQAD 328

QY 354 TIKVAV-----DRSCDIGNFEMLLRQPVVEPPFPFGSDLCVAGV--SVRRAR 398
DB 329 TLRVIDRDLNCDTGAPELIVKQPVTTSPASTPRSPCLLPMYSDTRAR 378

RESULT 8
AAB58909
ID AAB58909 standard; protein; 362 AA.
XX
```


CC cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
CC dermatological; immunosuppressive; antiinflammatory; antibacterial;
CC antiviral; antifungal; antithematic; antithyroid; and antianaemic. The
CC sequences can be used for determining the presence of or predisposition
CC to, or preventing or treating pathological conditions associated with an
CC ORFX-associated disorder. The nucleic acids can be used to express ORFX
CC proteins in gene therapy vectors. The proteins and nucleic acids may be
CC used to treat cancers, proliferative disorders, neurodegenerative
CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
CC storage, systemic lupus erythematosus, severe combined immunodeficiency
CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
CC cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to
CC enhance coagulation; to inhibit thrombosis; and as a contraceptive
XX
SQ Sequence 352 AA;

Query Match 33.5%; Score 956.5; DB 3; Length 352;
Best Local Similarity 54.2%; Pred. No. 2.2e-82;
Matches 179; Conservative 55; Mismatches 91; Indels 5; Gaps 2;

QY 74 DIDTSADAVEDLFEAWEDLTQYYSLVHGDAFISNRNYSFQOALLNRITSVNPQTDI 133
Db 2 DIDKLEAPLYLTPEGWSLFLQRYQVHVHGAELRLDVTQVQCEILQLOQAVVPQIDM 61

QY 134 DGLRNWIIPKAASGRDIVCMRDVEEILLEAAADHPLSRDNKVVQVQKIETPLLICDT 193
Db 62 EGRNWIWVKGAKSRGIMCMCHLEMLKLVNGNPVVMKDGKVVQVQKIETPLIFGT 121

QY 194 KFDIROWFLVDNPLTIWFKESYLRFSTQRFSLDKLSAHLNNAVQKLVNDVGRS 253
Db 122 KFDLQWFLVDNPLTIWFRDYSYRFTQPSLKNLDSVHLNNSIQHLENSCHRH 181

QY 254 PLLPAHNMWTSRFOEYLQQRGAVGWSVIVPSMKKAIHAAMKVAADHVEPRKNSFELY 313
Db 182 PLLPPDNMWSRQFQHLQEMGAPNAWSTIIVPGMKDVAIHALQTSODTVQCRKASFELY 241

QY 314 GADFVLGRDPRPWLIEINSSPTWHPSTPTAQLCAQVQEDTIKAVV----DRSCDIGNFE 369
Db 242 GADVFEGEDFQPLWLEINASPTWAPSTAVTARLCAGVQADTLRVVDRDLRNCDTGAFE 301

QY 370 LLWRQPVVEPPFSGDLGVAGV-SVRRAR 398
Db 302 LIYQPVVTTSPASTPRPSCLLPMSYDTRAR 331

RESULT 10
AAM39450
ID AAM39450 standard; protein; 352 AA.
XX
AC AAM39450;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 2595.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-brager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US034263.
XX
PR 23-DEC-1999; 99US-00471275.

PR 21-JAN-2000; 2000US-00488725.
PR 25-APR-2000; 2000US-00582317.
PR 20-JUN-2000; 2000US-00598042.
PR 19-JUL-2000; 2000US-00620312.
PR 03-AUG-2000; 2000US-00653450.
PR 14-SEP-2000; 2000US-00662191.
PR 19-OCT-2000; 2000US-00693036.
PR 29-NOV-2000; 2000US-00727344.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
PI Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI; 2001-442253/47.
DR N-PSDB; AAI58606.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders such
as central nervous system injuries.
XX
PS Example 4; SEQ ID NO 2595; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and the
CC encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-brager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders. Note: The sequence data for this patent did not form
CC part of the printed specification
XX
SQ Sequence 352 AA;

Query Match 33.5%; Score 956.5; DB 4; Length 352;
Best Local Similarity 54.2%; Pred. No. 2.2e-82;
Matches 179; Conservative 55; Mismatches 91; Indels 5; Gaps 2;

QY 74 DIDTSADAVEDLFEAWEDLTQYYSLVHGDAFISNRNYSFQOALLNRITSVNPQTDI 133
Db 2 DIDKLEAPLYLTPEGWSLFLQRYQVHVHGAELRLDVTQVQCEILQLOQAVVPQIDM 61

QY 134 DGLRNWIIPKAASGRDIVCMRDVEEILLEAAADHPLSRDNKVVQVQKIETPLLICDT 193
Db 62 EGRNWIWVKGAKSRGIMCMCHLEMLKLVNGNPVVMKDGKVVQVQKIETPLIFGT 121

QY 194 KFDIROWFLVDNPLTIWFKESYLRFSTQRFSLDKLSAHLNNAVQKLVNDVGRS 253
Db 122 KFDLQWFLVDNPLTIWFRDYSYRFTQPSLKNLDSVHLNNSIQHLENSCHRH 181

QY 254 PLLPAHNMWTSRFOEYLQQRGAVGWSVIVPSMKKAIHAAMKVAADHVEPRKNSFELY 313
Db 182 PLLPPDNMWSRQFQHLQEMGAPNAWSTIIVPGMKDVAIHALQTSODTVQCRKASFELY 241

QY 314 GADFVLGRDPRPWLIEINSSPTWHPSTPTAQLCAQVQEDTIKAVV----DRSCDIGNFE 369
Db 242 GADVFEGEDFQPLWLEINASPTWAPSTAVTARLCAGVQADTLRVVDRDLRNCDTGAFE 301

QY 370 LLWRQPVVEPPFSGDLGVAGV-SVRRAR 398
Db 302 LIYQPVVTTSPASTPRPSCLLPMSYDTRAR 331

RESULT 11
ADJ93457
ID ADJ93457 standard; protein; 352 AA.

XX The invention relates to human tumour-associated antigenic target (TAT)
CC polypeptides, and their related nucleic acids. The TAT polypeptides are
CC overexpressed in cancer tissues compared to normal tissues, and may thus
CC serve as effective targets for the diagnosis and treatment of cancer in
CC mammals. The invention also relates to nucleic acid and polypeptide
CC sequences at least 80% identical to the TAT nucleic acids and
CC polypeptides; expression vectors and host cells comprising a TAT nucleic
CC acid; an antibody specific for a TAT polypeptide; a peptide or organic
CC molecule which binds to a TAT polypeptide; fusion proteins comprising a
CC TAT polypeptide; and methods and compositions for the treatment or
CC diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,
CC antibodies, antagonists, binding molecules and compositions are useful
CC for diagnosing or treating a cell proliferative disorder associated with
CC increased TAT expression, particularly cancers such as breast cancer,
CC colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder
CC cancer, pancreatic cancer, cervical cancer, cancers of the central
CC nervous system, melanoma and leukaemia. TAT nucleic acids may further be
CC used as hybridisation probes, in chromosome and gene mapping, in
CC chromosome identification and in gene therapy. The present sequence
CC represents a TAT polypeptide of the invention
XX
SQ Sequence 352 AA;

Query Match 33.5%; Score 956.5; DB 8; Length 352;
Best Local Similarity 54.2%; Pred. No. 2.2e-82;
Matches 179; Conservative 55; Mismatches 91; Indels 5; Gaps 2;
QY 74 DITSDADVEDLTEAEWEDLTQOYSLVHGDAFISNSRNYFSQCQALLNRITSVNPQTDI 133
DB 2 DIDKDLAPLYLTPEGWSLFQRYQVHGAELRLDVTQVQCEDILQQLQAVVPQIDM 61
QY 134 DGLRNWIIPKAASGRDIVCMRVVEEILELAAADHPLSRDNKVVQKYIETPLLICDT 193
DB 62 EGDNRNIWVPGAKSRGIMCMCHLEMLKLVNGNPVVMKDGKVVQKYIERPLLIFFT 121
QY 194 KFDIROWFLVDNPLTIWFKESYLRFSTQFSLKLDLSAHLNNVAVOKYLNVDVGRS 253
DB 122 KFDLROWFLVDNPLTIWFKESYLRFSTQFSLKLDLSAHLNNVAVOKYLNVDVGRS 181
QY 254 PLLPAHNMWTSRFOEYLQGRGAVGVSVIYPSMKKAIHAAMKVAQDHVEPRKNSFELY 313
DB 182 PLLPPDNWSSQRFQAHLEMGAPNAWSTIIVFGMKDAVIHALQTSQDTVQCRKASFELY 241
QY 314 GADFVLGRDPRPWLIEINSSPTMHPSTFVTAQLCAQVQEDTIKAVV----DRSCDIGNFE 369
DB 242 GADFVFGEDFQPWLEINASPTMAPSTAVTARLCAGVQADTLRVVDRMLDRNCDTGAFE 301
QY 370 LLWRQPVVEPPFSGSLCVAGV-SVRRAR 398
DB 302 LIYKQPVTTSPASTPRPSCLLPMYSDTRAR 331

RESULT 13
ADM05524
ID ADM05524 standard; protein; 326 AA.
AC ADM05524;
XX
XX 20-MAY-2004 (first entry)
DE Human protein of the invention SEQ ID NO:4209.
XX human; gene therapy; diagnostic marker; pharmaceutical.
XX Homo sapiens.
XX EP1347046-A1.
XX 24-SEP-2003.
XX 12-APR-2002; 2002EP-00008400.
XX

PR 22-MAR-2002; 2002JP-00137785.
XX
XX (REAS-) RES ASSOC BIOTECHNOLOGY.
PA
PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
XX
XX WPI; 2003-723558/69.
DR N-PSDB; ADM03081.
XX
XX New polynucleotides and polypeptides are useful in gene therapy, for
PT developing a diagnostic marker or medicines for regulating their
PT expression and activity, or as a target of gene therapy.
XX
XX Claim 1; SEQ ID NO 4209; 305pp; English.
XX
XX The invention relates to a novel human polynucleotide and the encoded
CC polypeptide. A polynucleotide of the invention may have a use in gene
CC therapy. An oligonucleotide of the invention ADM06202-ADM06773 is useful
CC as a primer for synthesizing the polynucleotide or as a probe for
CC detecting the polynucleotide. The polynucleotides ADM0316-ADM03758 are
CC useful in gene therapy, for developing a diagnostic marker or medicines
CC for regulating their expression and activity, or as a target of gene
CC therapy. The proteins ADM03759-ADM06201 encoded by the polynucleotides
CC are useful as pharmaceutical agents. The present sequence represents a
XX protein sequence of the invention.
XX
SQ Sequence 326 AA;

Query Match 33.1%; Score 944.5; DB 7; Length 326;
Best Local Similarity 54.3%; Pred. No. 2.8e-81;
Matches 176; Conservative 55; Mismatches 86; Indels 7; Gaps 2;
QY 74 DITSDADVEDLTEAEWEDLTQOYSLVHGDAFISNSRNYFSQCQALLNRITSVNPQTDI 133
DB 2 DIDKDLAPLYLTPEGWSLFQRYQVHGAELRLDVTQVQCEDILQQLQAVVPQIDM 61
QY 134 DGLRNWIIPKAASGRDIVCMRVVEEILELAAADHPLSRDNKVVQKYIETPLLICDT 193
DB 62 EGDNRNIWVPGAKSRGIMCMCHLEMLKLVNGNPVVMKDGKVVQKYIERPLLIFFT 121
QY 194 KFDIROWFLVDNPLTIWFKESYLRFSTQFSLKLDLSAHLNNVAVOKYLNVDVGRS 253
DB 122 KFDLROWFLVDNPLTIWFKESYLRFSTQFSLKLDLSAHLNNVAVOKYLNVDVGRS 181
QY 254 PLLPAHNMWTSRFOEYLQGRGAVGVSVIYPSMKKAIHAAMKVAQDHVEPRKNSFELY 313
DB 182 PLLPPDNWSSQRFQAHLEMGAPNAWSTIIVFGMKDAVIHALQTSQDTVQCRKASFELY 241
QY 314 GADFVLGRDPRPWLIEINSSPTMHPSTFVTAQLCAQVQEDTIKAVV----DRSCDIGNFE 369
DB 242 GADFVFGEDFQPWLEINASPTMAPSTAVTARLCAGVQADTLRVVDRMLDRNCDTGAFE 301
QY 370 LLWRQ---PVVEPPFSGSLCVVA 390
DB 302 LIYKQGPAPNMQVSPERNAPLCVA 325

RESULT 14
ADJ93455
ID ADJ93455 standard; protein; 292 AA.
XX
XX ADJ93455;
XX
XX 06-MAY-2004 (first entry)
XX
XX Human HOTTTL protein sequence SeqID2.
DE
XX testis-specific tubulin tyrosine-ligase-like polypeptide;
XX BGS-42 polypeptide; cytostatic; respiratory-Gen; gastrointestinal-Gen;
XX neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;
XX osteopathic; nootropic; antiparkinsonian; antiarthritic; antiaesthatic;
XX

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 4, 2006, 09:58:51 ; Search time 4791 Seconds
(without alignments)
450.890 Million cell updates/sec

Title: US-10-635-977-2

Perfect score: 2854

Sequence: 1 MASSILKVVVSHQSCRSR.....LRGLTAEGALRPPPGKGS 541

Scoring table:

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 9263891 seqs, 1996499642 residues

Total number of hits satisfying chosen parameters: 18527782

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp
-O=/abs/ABSSWEB/spool/US10635977/runat_04042006_103546_10350/app_query_faeta.1
-DB=Published Applications NA New -QFMT=fastap -SUFFIX=p2n.rnpbn -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITRS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-MINLEN=2000000000 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-USER=US10635977 @CGN 1 1.986 @runat_04042006_103546_10350 -NCPU=6 -ICPU=3
-NO_WMAP -NEG_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications NA New:

1: /SID55/ptodata/2/pubpna/US08_NEW_PUB.seq:
2: /SID55/ptodata/2/pubpna/US06_NEW_PUB.seq:
3: /SID55/ptodata/2/pubpna/US07_NEW_PUB.seq:
4: /SID55/ptodata/2/pubpna/PCT_NEW_PUB.seq:
5: /SID55/ptodata/2/pubpna/US09_NEW_PUB.seq:
6: /SID55/ptodata/2/pubpna/US09_NEW_PUB.seq:
7: /SID55/ptodata/2/pubpna/US10_NEW_PUB.seq:
8: /SID55/ptodata/2/pubpna/US10_NEW_PUB.seq:
9: /SID55/ptodata/2/pubpna/US10_NEW_PUB.seq:
10: /SID55/ptodata/2/pubpna/US10_NEW_PUB.seq:
11: /SID55/ptodata/2/pubpna/US11_NEW_PUB.seq:
12: /SID55/ptodata/2/pubpna/US11_NEW_PUB.seq:
13: /SID55/ptodata/2/pubpna/US11_NEW_PUB.seq:
14: /SID55/ptodata/2/pubpna/US11_NEW_PUB.seq:
15: /SID55/ptodata/2/pubpna/US60_NEW_PUB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	922	32.3	1684	8 US-10-955-054A-43	Sequence 43, Appli
2	357	12.5	666	6 US-09-925-065A-953009	Sequence 953009,

3	302	10.6	2624	11	US-11-072-512-1688	Sequence 1688, Ap
4	271.5	9.5	2074	11	US-11-072-512-1401	Sequence 1401, Ap
5	210	7.4	2111	11	US-11-072-512-1651	Sequence 1651, Ap
6	144.5	5.1	2250	9	US-10-932-182A-167	Sequence 167, App
7	144.5	5.1	2250	9	US-10-932-182A-167	Sequence 167, App
8	141	4.9	1319	8	US-10-750-185-41745	Sequence 41745, A
9	141	4.9	1319	8	US-10-750-623-41745	Sequence 41745, A
10	136.5	4.8	47460	14	US-11-124-368A-2877	Sequence 2877, Ap
11	132.5	4.6	10373	8	US-10-821-234-64	Sequence 64, Appli
12	127	4.4	2534	14	US-11-219-995-1	Sequence 1, Appli
13	126.5	4.4	2754	9	US-10-330-773-662	Sequence 662, Appli
14	126	4.4	5354	14	US-11-136-527-1889	Sequence 1889, Ap
15	126	4.4	14248	8	US-10-995-561-13381	Sequence 13381, A
16	126	4.4	96128	8	US-10-995-561-13197	Sequence 13197, A
17	123.5	4.3	3482	14	US-11-064-246-9	Sequence 9, Appli
18	123.5	4.3	3482	14	US-11-064-246-11	Sequence 11, Appli
19	122.5	4.3	997	10	US-10-301-480-609833	Sequence 609833,
20	122.5	4.3	997	10	US-10-301-480-1223242	Sequence 1223242,
21	122.5	4.3	5382	9	US-10-501-035-116	Sequence 116, App
22	121	4.2	6409	9	US-10-501-035-158	Sequence 158, App
23	121	4.2	15457	14	US-11-136-527-2809	Sequence 2809, Ap
24	120	4.2	551	6	US-09-925-065A-360042	Sequence 360042,
25	120	4.2	551	6	US-09-925-065A-360043	Sequence 360043,
26	119	4.2	1724	6	US-09-925-065A-552157	Sequence 552157,
27	119	4.2	1724	6	US-09-925-065A-552158	Sequence 552158,
28	119	4.2	1724	10	US-10-301-480-530107	Sequence 530107,
29	119	4.2	1724	10	US-10-301-480-530108	Sequence 530108,
30	119	4.2	1724	10	US-10-301-480-1143516	Sequence 1143516,
31	119	4.2	1724	10	US-10-301-480-1143517	Sequence 1143517,
32	118	4.1	1724	6	US-09-925-065A-552159	Sequence 552159,
33	118	4.1	1724	10	US-10-301-480-530109	Sequence 530109,
34	118	4.1	1724	10	US-10-301-480-1143518	Sequence 1143518,
35	117.5	4.1	2980	11	US-11-072-512-353	Sequence 353, App
36	117.5	4.1	3261	8	US-10-624-932-7	Sequence 7, Appli
37	117.5	4.1	3465	8	US-10-624-932-5	Sequence 5, Appli
38	117.5	4.1	3734	8	US-10-131-826A-147	Sequence 147, App
39	117.5	4.1	3734	9	US-10-973-115B-147	Sequence 147, App
40	117	4.1	1263	9	US-10-016-686-14	Sequence 14, Appl
41	117	4.1	1862	8	US-10-750-185-58437	Sequence 58437, A
42	117	4.1	1862	8	US-10-750-623-58437	Sequence 58437, A
43	117	4.1	2430	9	US-10-643-457-3	Sequence 3, Appli
44	116.5	4.1	1488	14	US-11-219-995-3	Sequence 3, Appli
45	116.5	4.1	2149	11	US-11-096-568A-14615	Sequence 14615, A

ALIGNMENTS

RESULT 1

US-10-955-054A-43
; Sequence 43, Application US/10955054A
; Publication No. US20050266420A1
; GENERAL INFORMATION:
; APPLICANT: PUSZTAI, LAJOS
; APPLICANT: SYMMANS, W. FRASER
; APPLICANT: HESS, KENNETH R.
; APPLICANT: AYERS, MARK
; APPLICANT: STEC, JAMES
; TITLE OF INVENTION: MULTIGENE PREDICTORS OF RESPONSE TO CHEMOTHERAPY
; FILE REFERENCE: UTXC:880US
; CURRENT APPLICATION NUMBER: US/10/955.054A
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 195
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 43
; LENGTH: 1684
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-955-054A-43

Alignment Scores:

Pred. No.:

Score:

Percent Similarity:

2.37e-60

922.00

55.5%

Length:

Matches:

Conservative: 68

Best Local Similarity: 41.9%	Mismatches: 128		
Query Match: 32.3%	Indels: 96		
DB: 8	Gaps: 12		
US-10-635-977-2 (1-541) x US-10-955-054A-43 (1-1684)			
Qy	73 GluAepIleaspThrSerAlaaspAlaValGluAaspLeuThrGluAlaGluTrpGluAasp 92		
Db	161 AAGGATCTGATGACTCAGCGATGGGTGACGTGACGTGACACCACTGAGGATGAGGATGAAGAT 220		
Qy	93 LeuThrGlnGlnTyr----- 97		
Db	221 GAGGACGAGGAGTTCAGCCCATCAGACTGTTGCGACTTCGATGATTTACTGAATTTGAT 280		
Qy	98 -----TyrSerLeuVal-HisGlyAaspAlaPheIleSerAsnSerArgAs 112		
Db	281 GACCTAGATGGAACACATGCTCTGATGCGAAGGG--GCAGAACTCAGGCACCTCGACAC 337		
Qy	112 nTyrPheSerGlnCysGlnAlaLeuLeuAsnArgIleThrSerValAsnProGlnThrAs 132		
Db	338 TCAGGTCACGCGTGTGAGGACATCTCGCAGCAGCTGTCAGGCGGTGTACCCAGATAGA 397		
Qy	132 pIleAaspGlyLeuArgAsnIleTrpIleIleIysProAlaAlaLysSerArgGlyArgAs 152		
Db	398 CATGGAAGGGGATCGCAACTTCTGATCTGTAAGCAGGAGCAAGTCCCGCGACGAGG 457		
Qy	152 pIleValCysMetAspArgValGluGluIleLeuGluLeuAlaAlaAspHisProLe 172		
Db	458 CATCATGTGATGACCACTCGAGGAGATGCTGAAGCTGTGTAAGCGCAACCCCGTGGT 517		
Qy	172 uSerArgAspAsnLysTrpValValGlnLysTyrIleGluThrProLeuLeuIleCysAs 192		
Db	518 GATGAAGGACGGCAAGTGGTGTGCAGAAGTATATTGAGCGGCCCTCTCATCTTTGG 577		
Qy	192 pThrLysPheAaspIleArgGlnTrpPheLeuValThrAspTrpAsnProLeuThrIleTr 212		
Db	578 CACCAAGTTTGACCTCAGACAGTGGTTCCTGGTAACTGACTGGAACCCACCTTACCGGTG 637		
Qy	212 pPheTyrLysGluSerTyrLeuArgPheSerThrGlnArgPheSerLeuAspLysLeuAs 232		
Db	638 GTTCTACCGGACAGCTATATCCGCTTTCCACGACGCCCTTCTCTCTGAAGAACCTGGA 697		
Qy	232 pSerAlaIleHisLeuCysAsnAsnAlaValGlnLysTyrLeuLysAsnAspValGlyAr 252		
Db	698 CAACCTCAGTCACCTGTGCACAACTCCATCCAGAGACCTGGAGAAGCTCATGCGCATCG 757		
Qy	252 gSerProLeuLeuProAlaHisAsnMetTrpThrSerThrArgPheGlnGluTyrLeuGl 272		
Db	758 GCATCCACTGCTTCCGCGACAGCAACATGTGTGTAGCCAGAGGTTCCAGGCCACCTGCA 817		
Qy	272 nArgGlnGlyArgGlyAlaValTrpGlySerValIleTyrProSerMetLysLeuAlaI 292		
Db	818 GGATGGGTGGCCCCAAATGCTTGGTCCACCATCATCTGCTGCTGGCATGAAGATGCTGT 877		
Qy	292 eAlaHisAlaMetLysValAlaGlnAspHisValGluProArgLysAsnSerPheGluLe 312		
Db	878 GATCAGCACTTCAGACCTCCAGACACCGTGCAGTGTGGAAGGCCAGCTTTGAGCT 937		
Qy	312 uTyrGlyAlaAspPheValLeuGlyArgAspPheArgProTrpLeuIleGluLeuAsnSe 332		
Db	938 CTATGGCGCTGACTTCGTGTGCGGGGAGGACTTCCAGCCCTGGCTGATTGAGATCAACGC 997		
Qy	332 rSerProThrMetHisProSerThrProValThrAlaGlnLeuCysAlaGlnValGlnGl 352		
Db	998 CAGGCCCAAGATGACCCCTCCACAGCAGTCACTGCCCGGCTCTGTGCTGGCGTGCAGC 1057		
Qy	352 uAspThrIleLysValAlaVal-----AspArgSerCysAspIleGlyAsnPh 368		
Db	1058 TGACACCTGCGCGTGTGTCATTCAGCGGATGCTGAGCGCACTGTGACACAGGACCTT 1117		
Qy	368 eGluLeuLeuTrpArgGlnProValValGluProPro----- 380		
Db	1118 TGAGCTCATCTATAGACAGCCGCTCACCACTTCCACAGCGCTCCACACCAAGGCCAGCTG 1177		
Qy	381 -----ProPhe-----SerGlySerAspLeuCysValAlaGl 391		
Db	1178 CCTTCTCCCATGTACTCCGACACACAGGCCAGGCTCTCAGACGACGACGACGACGCTG 1237		
Qy	391 yValSerValArgAlaArgGlnValLeuProValCysAsnLeuLysAlaSerAl 411		
Db	1238 GTGGGACATAAGCCCTGTGCACACAGCAAGGCGCTTG-AGGACTCTACCCACGGCTAA 1296		
Qy	411 aSerLeuLeuAspAlaGlnPro-----LeuLysAl 421		
Db	1297 GGTCTTCATTTCCTCCACCGAACCTTGATTTCAGGTGGCACCAGCATCTCTGAAGCC 1356		
Qy	421 aArgGlyProSerAlaMetProAspProAlaGlnGlyProProSerProAlaLeuGlnAr 441		
Db	1357 AAGAAA-GGCTCTCTCTTCTGTGCTCCGAGGCCGCCAGCTGGAAGTGCCTTG----- 1410		
Qy	441 gAspLeuGlyLeuLysGluLysGlyLeuProLeuAlaLeuLeuAlaProLeuArgGl 461		
Db	1411 -----TTGCCCTCTGCCCTTT----- 1425		
Qy	461 yAlaAlaGluSerGlyGlyAlaAlaGlnProThrArg---ThrLysAlaAlaGlyLysVa 480		
Db	1426 -----GAAGTCGGACAATTCCTAGCACCTGTGCGAAGGTCAAGGCCAAAGCAAATT 1478		
Qy	480 lGlu-----LeuProAlaCysProCysArgHisVa 490		
Db	1479 CAAGGCCAGACTGTGCACAAACCCAGGCGCTGAGGCGCTGCC-----CATGA 1523		
Qy	490 lAspSerGlnAlaProAsnThrGlyValProValAlaGlnProAlaLysSerTrp 508		
Db	1524 AGAGGCTGAGCCCCCTGAAACC-----CTGCCCTTGTGTGG 1560		
RESULT 2			
US-09-925-065A-953009/c			
; Sequence 953009, Application US/09925065A			
; Publication No. US20040181048A1			
; GENERAL INFORMATION:			
; APPLICANT: Wang, David G.			
; TITLE OF INVENTION: Identification and Mapping of Single			
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome			
; FILE REFERENCE: 108827.135			
; CURRENT APPLICATION NUMBER: US/09/925,065A			
; PRIORITY FILING DATE: 2001-08-08			
; PRIOR APPLICATION NUMBER: US 60/243,096			
; PRIORITY FILING DATE: 2000-10-24			
; PRIOR APPLICATION NUMBER: US 60/252,147			
; PRIORITY FILING DATE: 2000-11-20			
; PRIOR APPLICATION NUMBER: US 60/250,092			
; PRIORITY FILING DATE: 2000-11-30			
; PRIOR APPLICATION NUMBER: US 60/261,766			
; PRIORITY FILING DATE: 2001-01-16			
; PRIOR APPLICATION NUMBER: US 60/289,846			
; PRIORITY FILING DATE: 2001-05-09			
; NUMBER OF SEQ ID NOS: 957086			
; SOFTWARE: FastSeq for Windows Version 4.0			
; SEQ ID NO 953009			
; LENGTH: 666			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
US-09-925-065A-953009			
Alignment Scores:			
Pred. No.:	6,16e-18	Length:	666
Score:	357.00	Matches:	106
Percent Similarity:	46.4%	Conservative:	22
Best Local Similarity:	38.4%	Mismatches:	61
Query Match:	12.5%	Indels:	88
DB:	6	Gaps:	14
US-10-635-977-2 (1-541) x US-09-925-065A-953009 (1-666)			
Qy	171 ProLeuSerArgAspAsnLysTrpValValGlnLysTyrIleGluThrProLeuLeu 190		

```
Db 666 CCTCTTTCCAGGACAAACAGTGGTGGTCCAGAGTACATCAGACGCGCTGCTCATC 607
Qy 191 CysAspThrLysPheAspIleArgGlnTrpPheLeuValThrAspTrpAsnProLeuThr 210
Db 606 TGTGACCAAGTTCGACATCAGACAGTGTCTCGTCACGACTGGAAACCCCTGACC 547
Qy 211 IleTrpPheTyrLysGluSerTyrLeuArgPheSerThrGlnArgPheSerLeuAspLys 230
Db 546 ATCTGGTTCTACAAGGAGAGTTACTTTCGGTTCCTCAACTCAGCGCTTCTCCCTGACAG 487
Qy 231 LeuAspSerAlaIleHis-----Leu 237
Db 486 CTGGACAGTCACTGGTGGGCGGACCTGGACACCTCGGCGCAGGGAATGGCTCTGCC 427
Qy 238 CysAsnAsnAlaValGlnLysTyrLeuLysAsnAspValGlyArgSerProLeuLeuPro 257
Db 426 TGCATGCTGGAGCAGCCGCA---GTGAGACGACGACGGGCGGCGCCCTCATCTGCCA 370
Qy 258 AlaHisAsnMetTrpThrSerThrArgPheGlnGluTyrLeuGlnArgGlnGlyArgGly 277
Db 369 CCT-----GCCACACGAGAGGCACAGAT-----GGGGAACCT 337
Qy 278 AlaValTrpGlySerValIleTyrProSerMetLysLysAlaIleAlaHisAlaMetLys 297
Db 336 GAGGCTTGGGAGGCTGATGATCTCCCGGAGGACAG----- 301
Qy 298 ValAlaGlnAspHisValGluProArgLysAsnSerPheGluLeuTyrGlyAlaAspPhe 317
Db 300 ---GCCACGAGCCAC----- 289
Qy 318 ValLeuGlyArgAspPheArgProTrpLeuIleAsnSerSerProThrMethHis 337
Db 288 -----CGGCT-----TCAAAGCCAGCG----- 271
Qy 338 ProSerThrProValThrAlaGlnLeuCysAlaGlnValGlnGluAspThrIleLysVal 357
Db 270 CCTTACACCCACGCTCTGTGCATCATCTCGGAAATCTTAAAGGACCACTT---TTG 214
Qy 358 AlaValAspArgSerCysAspIleGlyAsnPheGluLeuLeuTrpArgGlnProValVal 377
Db 213 CCTGTGCCAGGGG-----ATTGGA-----ATGGTGTGG-----GCA 181
Qy 378 GluProProPheSerGlySerAspLeuCysValAlaGlyValSerValArgAla 397
Db 180 GCACCTGGGCTCG----- 158
Qy 398 ArgArgGlnValLeuProValCysAsnLeuLysAlaSerAlaSerLeuLeuAspAlaGln 417
Db 157 AGCGCGCTGTGGGTGCTGTGGAGGCTCAGATATCCGCAAGTGTCTGCC----- 104
Qy 418 ProLeuLysAlaArgGlyProSerAlaMetProAspProAlaGlnGly 433
Db 103 -----AGGTGGATGGACCCCACTCACCCTGGTCCCTCCGCTGGGCG 62
```

RESULT 3

```
US-11-072-512-1688
; Sequence 1688, Application US/11072512
; Publication No. US20060029945A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
```

Alignment Scores: 3.43e-13 Length: 2624
Pred. No.: 302.00 Matches: 132
Score: 37.7% Conservative: 76
Percent Similarity: 24.0% Mismatches: 200
Best Local Similarity: 10.6% Indels: 144
Query Match: 11 Gaps: 23
DB: US-10-635-977-2 (1-541) x US-11-072-512-1688 (1-2624)

Qy 82 ValGluAspLeuThrGluAlaGluTrpGlu-----GlnGlnTyr 97
Db 447 GTGGAAGTGAAGGAGGAGGAGTGGATTTCTACTGGTGTGACGTGAGTGGCTCGCG 506
Qy 92 -----AspLeuThr-----GlnGlnTyr 97
Db 507 GAGAACTTCGACACACCTACATGGATGAACATGTGCGGATCAGTCACTTCGGAACAC 566
Qy 98 TyrSerLeuValHisGlyAspAlaPheIleSerAsnSerArgAsnTyrPheSerGlnCys 117
Db 567 TATGAGTGTGACCCGGAAGACTACATGTTGAAGAACCTGAAGCGGTTCGGAAGCAGCTG 626
Qy 118 GlnAlaLeuLeuAsnArgIleThrSerValAsn-----ProGlnThr----- 131
Db 627 GAGCGTGAGCGAGGAAAGCTGCGAGCAGCAAGTGTGACTTCTTCCCAAAACCTTTGAG 686
Qy 132 -----AspIleAspGlyLeuArgAsn-----IleTrp 140
Db 687 ATGCTTTGCGAGTACCACCTGTTTGTAGAGGAGTTTCGCAAAACCCAGGAATCACCTGG 746
Qy 141 IleIleLysProAlaAlaLysSerArgGlyArgAspIleValCysMetAspArgValGlu 160
Db 747 ATCATGAAGCCTGTAGCCCGCTCTCAAGGGAAAGGCATCTTCTCTTCCGTAGGCTGAAG 806
Qy 161 GluIleLeuGluLeuAlaAlaAspHisProLeuSerArgAspAsnLys----- 177
Db 807 GACATCTGGAGTGG-----AGGAAGGACACAAAGAGCTCTGACGACCAAGATGATATT 863
Qy 178 -----TrpValValGlnLysTyrIleGluThrProLeuLeuIleCysAspThr 193
Db 864 CCCGTGGAGAACTATGTGGCTCAGCGTTACATTCGAAATCTTACCTGATAGGAGGCCGC 923
Qy 194 LysPheAspIleArgGlnTrpPheLeuValThrAspTrpAsnProLeuThrIleTrpPhe 213
Db 924 AAGTTTGCACCTGCGTGTCTATGTGTCTGCTGATGTCG-----GTG 962
Qy 214 TyrLysGluSerTyrLeuArgPheSerThrGlnArgPheSerLeuLeuAspLysLeuAspSer 233
Db 963 TTTGCTGAATCCCTGCTGTGCTGGGCACAGGAGA-----CAG 1001
Qy 234 AlaIleHisLeuCysAsnAsnAlaValGlnLysTyrLeuLysAsnAspValGlyArgSer 253

```
Db 1002 GATGTTACCTCACCACGCTGGCTGTGCAAAA----- 1034
Qy 254 ProLeuLeuProAlaHisAsn-----MetTrpThrSerThrArgPheGln 268
Db 1035 ---ACATCTCCGAGTACCACCCAAAGAGGGCTGCAAGTGAGCGCTGCAGCGCTTCCGG 1091
Qy 269 GluTyLeu---GlnArgGlnGlyArgGlyAlaValTrpGlySerValIleTyProSer 287
Db 1092 CAGTACCTCGGCTCCAAACACGCGGCCGAGCAGTG-----GAGACATCTTTCAGGGAC 1145
Qy 288 MetLysLysAlaIleAlaHisAlaMetLysValAlaGlnAspHisValGluProArgLys 307
Db 1146 ATGCACAACATCTTTGTCAAAAGCCTGCAGAGTGTGCAGAGGTGATCATCATGACACAAG 1205
Qy 308 AsnSerPheGluLeuTyArgAlaAlaAspPheValLeuGlyArgAspPheArgProTrpLeu 327
Db 1206 CACTGCTTCGAGCTGTACGGCTATGACATCTCATCGACCGAGCACCTCAAGCGCTGGCTC 1265
Qy 328 IleGluIleAsnSerSerProThrMetHisProSerThrProValThrAlaGlnLeuCys 347
Db 1266 CTGAGGCTCAATCGTCCCATCCTGACAGCGCAGCGACGCGAAGAACTATGAGCTCAAG 1325
Qy 348 AlaGlnValGlnGluAspThrIleLysValAla-----ValAspArg 361
Db 1326 ACCTGCTCTCGAAGACACCTTCATGTTGTGCATGGAAGCGAGGCTCACGGGAAGG 1385
Qy 362 SerCysAspIleGlyAsnPheGluLeuLeuTrpArgGlnProValValGlu----- 378
Db 1386 GAGAAGCGAGTCCGGGCTTTGACCTCATGTGAATGATGGCCCTGTTAGCAGAGAGGAG 1445
Qy 379 -----ProProProPheSerGlySerAspLeuCysValAlaGlyValSerValArg 396
Db 1446 GGGGCTCTCGA-CCTGTGCGGAATGGGAACTTTGTGACCAACACACATCTCGGCTCG 1504
Qy 397 AlaArgArg----- 399
Db 1505 CAACGTCGGAAGAAACAACTGAGGCGAGCTCTCTGCTCCTTCAAGTTTCAGAAAGC 1564
Qy 400 ---GlnValProValCysAsnLeuLysAlaSerAlaSerLeuLeuAspAlaGlnPro 418
Db 1565 TTCAGTGTGATCCCGAGCTCCAGGAGGAATAACAGCCTTAGC-----AGGCT 1612
Qy 419 Leu-----LysAlaArgGlyProSerAlaMetProAspPro 430
Db 1613 TGTGTTTGTGAGTTGGGAGCTGAGCCAGAAACCGGAAACCTCGGCTCATGGCGAGACA 1672
Qy 431 AlaGlnGlyProProSerProAlaLeuGlnArgAspLeuGlyLeuLysGluLysGly 450
Db 1673 GCTTCGGGAGTTGAGTGTCAAGCCCGAAGGAAGACTTT---TGGAAGGAAATACTGGC 1729
Qy 451 LeuProLeuAlaLeuLeuAlaProLeuArgGlyAlaAlaGlnSerGlyAlaAlaGln 470
Db 1730 AAATCCAAAGAGTGAACCGCGACTT---TAGGAAGCCAGAAAGCCAGGCTACTCTAGA 1786
Qy 471 ProThrArgThrLysAlaAlaGlyLysValGluLeuProAlaCysProCysArgHisVal 490
Db 1787 CCTTCTGCTTCAGAGCATGAGGACTTCTCCCACTCATCCCTGTACTCAGGCGACACAA 1846
Qy 491 AspSerGlnAlaProAsnThrGlyValProValAlaGlnProAlaLysSerTrpAspPro 510
Db 1847 CCCAGACAG-----GTCCAGGCAATGCGACTTCTTT 1873
Qy 511 AsnGlnLeuAsnAlaHisProLeuGluProValLeuArgGlyLeuLysThrAlaGluGly 530
Db 1874 GGTTCCCTA-----TTTCCAGAAGGAATGCGATTGCTTTTTT 1912
Qy 531 AlaLeuArgProPro-----ProGlyGlyLysGly 540
Db 1913 GGCCAGACACCCCTGCGCAGGGAAGGGG 1945
```

RESULT 4

US-11-072-512-1401

; Sequence 1401, Application US/11072512

```
; Publication No. US20060029945A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOTYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: 084335-0191
; CURRENT APPLICATION NUMBER: US/11/072,512
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: US 60/350,978
; PRIOR FILING DATE: 2002-01-25
; PRIOR APPLICATION NUMBER: JP 2001-379298
; PRIOR FILING DATE: 2001-11-05
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1401
; LENGTH: 2074
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-11-072-512-1401

Alignment Scores:
Pred. No.: 5 59e-11 Length: 2074
Score: 271.50 Matches: 107
Percent Similarity: 36.6% Conservative: 59
Best Local Similarity: 23.6% Mismatches: 149
Query Match: 9.5% Indels: 139
DB: 11 Gaps: 15

US-10-635-977-2 (1-541) x US-11-072-512-1401 (1-2074)
Qy 2 AlaSerSerIleLeuLysTrpValValSerHisGlnSerCysSerArgSerSerArgSer 21
Db 14 GCCTCCGCGCATTTACGCTGGCGCTGCAGC-----TGCTGCACAGAGACCCACAGAGGC 67
Qy 22 LysProArgAspGlnArgGluAlaGlySerSerAspLeuSerArgGln----- 39
Db 68 CACCCTCGGAACACGCGCCCAATGAGAGGGCGGAGCCTGTGTCTCCACACAAAGCCAG 127
Qy 40 -----AspAla 41
Db 128 GCGCTGGATCTTTGAGAACCCACCACCCAGCCCTTACCTTAACATTCATCCGAGGCA 187
Qy 42 GluAsnAlaGluAlaLysLeuArgGlyLeuProGlyGlnLeuValAspIleAlaCysLys 61
Db 188 AACACACTGACGAGCGCGCTGCAGGCGCTGGAGCAAGGCTACAGGAAGCAGGTGTTTCC 247
Qy 62 Val----- 62
Db 248 ATCCCTCCAGCGAGCGCCGCCAACACCAACTGGAGAGAGAGAAAAAACCTCATTTG 307
Qy 63 -----CysGlnAlaTyLeuGlyGlnLeuGluHisGluAspIleAsp 76
Db 308 ATGGCGGAAGATGAACCTTCAGGGGCCCTCTTGAAGCGCTGGTTTTCGGGTGACGAG 367
Qy 77 ThrSerAlaAspAlaValGluAsp-----LeuThrGluAlaGluTrpGluAspLeuThrGln 95
Db 368 ACCACCCCGGCTGTGTGTGCAAAAGTGTCTCTCTGGAGAGGGGGTGGAAATAGTTTGA 427
```



```
QY 96 GlnTyr-----TyrSerLeuValHisGlyAspAlaPheIleSerAsnSer 110
Db 428 CAGGAGCAGACGCGGAGGACTGGAACCTGTACTG-GAGGACATCTCTTCGGAATGAC 486
QY 111 ArgAsnTyrPheSerGlnCys-----
Db 487 CGAACA-----CAACAGTGTAAACCGTGGCGAGCAGCTAAACCCACCTGGGAACCA 540
QY 118 GlnAlaLeuAsnArgIleThrSerValAsnPro-----
Db 541 CAAGCTTACCAGGAAGACTGTTTGGCCAAACACCTGAGCAGACATGAGGAGGTGTATGG 600
QY 129 -----
Db 601 CACTTCCCTGTACCAGTTCCATCCCTGACGTTGCTCATGCCCAATGACTATACCAAGTT 660
QY 130 -----GlnThrAsp-IleAspGlyLeuArgAsnIleTrpIleIrl 142
Db 661 CGTGGCTGATACTTTCAGGAGGAGGAGATGCTGGGCACCAAGCATAGCTATTGGATTGG 720
QY 142 elysProAlaAlaLysSerArgGlyArgAspIleValCysMetAspArgValGluGluIrl 162
Db 721 CAAGCTGCTGAGTTATCTCGTGGGAGGGGATACTAATTTTCAGTGCATTTAAAGACTT 780
QY 162 eLeuGluLeuAlaAlaAspHisProLeuSerArgAspAsnLysTrpValGlnLny 182
Db 781 CATC-----TTTGATGATATGTACATAGTCAGAA 810
QY 182 sTyrIleGluThrProLeuLeuIleCysAspThrLysPheAspIleArgGlnTrpPheIe 202
Db 811 ATATATCTCCAATCTTTACTTTATTTGGCAGATATAAATGTGATCTCCGCATCTATGTTG 870
QY 202 uValThrAspTrpAsnProLeuThrIleTrpPheTyrLysGluSerTyrLeuArgPheSe 222
Db 871 TGTACTGCTTAAAGCCTTTCACCATTTATGTTTATCAGGAAGGGTGGTTCGTTTGC 930
QY 222 rThrGlnArgPheSerLeuAspLysLeu---AspSerAlaIleHisLeuCysAsnAl 241
Db 931 CACGGAAGAGTTTGACCTCAGTAATTTGCAAAACAATATATGCCCATTTGACCAACAGCAG 990
QY 241 aValGlnLys-----TyrLeuLysAsnAspValGlyArgSerProIe 255
Db 991 CATCAATAAATCCGGGGCTCTTATGGAAGATCAAGAAGTGTATGTGTCATGGT----- 1045
QY 255 uLeuProAlaHisAsnMetTrpThrSerThrArgPheGlnGluTyrLeuGlnArgGlnG 275
Db 1046 -----TGTAATGGACGCTCAGCAGATTTTTTCTTACCTTCGTAGC----- 1087
QY 275 yArgGlyAlaValTrp-----GlySerValIleTyrProSerMetLysLysAlaIrl 292
Db 1088 -----TGGGATGTGGACGATCTGCTTTTGTGGAAGAAATCCACCGCATGGT 1134
QY 292 eAlaHisAlaMetLysValAlaGlnAspHisValGluProArgLysAsnSerPheGluIe 312
Db 1135 TATTCTCACCATTCTCGCATTCGACCATCTGCTCCCTTTGCTGCAATATGCTTTGAGCT 1194
QY 312 uTyrGlyAlaAspPheValLeuGlyArgPheArgProTrpLeuIleGluIleAsnSe 332
Db 1195 CTTTGGGTTGATATTTTGTATGATGATGACAACTTGAAACCATGGCTTTTAGAGGTCACTA 1254
QY 332 rSerPro-----ThrMetHisProSerThrProVal 342
Db 1255 CAGCCACGCTTGACCTTGGATTGTTCAACAGATGTG 1291
```

RESULT 5

```
US-11-072-512-1651
; Sequence 1651, Application US/11072512
; Publication No. US20060029945A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
```

```
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YUKI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: 084335-0191
; CURRENT APPLICATION NUMBER: US/11/072,512
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: US 60/350,978
; PRIOR FILING DATE: 2002-01-25
; PRIOR APPLICATION NUMBER: JP 2001-379298
; PRIOR FILING DATE: 2001-11-05
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1651
; LENGTH: 2111
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-072-512-1651
```

```
Alignment Scores:
Pred. No.: 2,6e-06 Length: 2111
Score: 210.00 Matches: 146
Percent Similarity: 32.7% Conservative: 53
Best Local Similarity: 24.0% Mismatches: 202
Query Match: 7.4% Indels: 210
DB: 11 Gaps: 31
US-10-635-977-2 (1-541) x US-11-072-512-1651 (1-2111)
```

```
QY 6 LeuLysTrpValValSerHisGlnSerCysSerArgSerArgSerLysProArgAsp 25
Db 532 CTGAAGTGG-----TGTCAGGTCAAGACCGCAGACAGCTACGCGCAGC 573
QY 26 GlnArgGluGluAlaGlySerSerAspLeuSerArgGlnAspAla----- 41
Db 574 TTCGGGGAAGG-AGAGCAGCTGCTGTACCAGCTTCCCAACAACAAAGCTCCTCACCACCAA 632
QY 42 GluAsnAlaGlu-AlaLysLeuArgGlyLeuProGlyGlnLeuValAspIleAlaCysLy 61
Db 633 GATCGGGCTGCTCAGCACCTTTCGGGA---CGGGCAGGGCCATGACGAGGCCAGCA 689
QY 61 sValCysGlnAlaTyrLeuGlyGlnLeuGluHisGluAspIleAspThrSerAlaAspAl 81
Db 690 GGTGCGGGGGGTCCAGGCCAGGCTGGAAGAAG-----GACGAGCAGCGCGCGC 740
QY 81 aValGluAspLeuThrGluAlaGluTrpGluAspLeuThrGlnGlnTyrTyrSerLeuVa 101
Db 741 CCTGGAGGACCTC-----CCGTGG----- 759
QY 101 lHisGlyAspAlaPheIleSerAsnSerArgAsnTyrPheSerGlnCysGlnAlaLeuLe 121
Db 760 -----ACAAGCCAGGATACCTCAGGCCACAGAGGGTCTGTG--- 795
QY 121 uAsnArgIleThrSerValAsnProGlnThr-----AspIleAspGlyLeuArg--- 137
Db 796 -----AGAATGAAGAGAGTTTTTCCAGAGACCTTACCGCTGGACCTCAAAACAGAGAGA 851
QY 138 -----AsnIleTrpIleIleLysProAlaAlaLysSe 148
Db 852 GGCCTTTTTCACCTGTTTGTATGAACCCAGATATGGATCTGCAAGCCCAACAGCTCCAA 911
```


1564	Db	AATGCAGANTTCATGGATAATAACAAGTAATAATCTCCAAATGCGCCACTTTATCATATA	1623
181	Qy	GlnLysTyrIleGluThrProLeuLeuIleCys-----AspThrLysPheAspIle	197
1624	Db	CAAGATAATTCAACAGCGCCATTATTGCTATCATCCATGATCAATAAGAAAGTTTCATATA	1683
198	Qy	ArgGlnIrrPheLeuValThrAspTrpAsnProLeuThrIleTrpPheTyrLysGluSer	217
1684	Db	AGGTGTTAT-----GTTGTTGTAAGGGGATTTACAGTATTCGTATACGATGAGTATG	1737
218	Qy	TyrLeuArgPheSerThrGlnArgPhe-----SerLeu	228
1738	Db	CTAGCACTATTTCGTGCTAAACCGTTTGTCCCTCTAGATCCTTAACACATATTCGGTAACA	1797
229	Qy	AspLysLeuAspSerAlaIleHisLeuCysAsnAsnAlaValGlnLysTyrLeuLysAsn	248
1798	Db	GATTTGAAAGCATTTGGAGTGCCATCTGACTAATAACATGCTTGCAAAGC---AAGAAAGAAA	1854
249	Qy	AspValGlyArgSerProLeuLeuProAlaHisAsnMetTrpThrSerThrArgPheGln	268
1855	Db	GATAAAGATTCTCTGTTTG-----GAGTTTGAT	1884
269	Qy	GluTyrLeuGlnArgGlnGlyArgGlyAlaValTrpGlySerValIleTyrProSerMet	288
1885	Db	TCCCTAGACGAA-----ATTCCAAACGAG	1908
289	Qy	LysIys-----AlaIleAlaHisAla	295
1909	Db	AAAAAGTTGAAATTAAGGAACAGATTCATTGTCATAACAAATGACGCTTTCTTAGCGGCA	1968
296	Qy	MethysValAlaGlnAspHisValGluProArgLysAsnSerPheGluLeuTyrGlyAla	315
1969	Db	GTAAATGTAACAGCGCTAAACTTCCAACCGTTACCAAAATGCGTTTGAGACCTATGTTGTA	2028
316	Qy	AspPheValLeuGlyArgAspPheArgProTrpLeuIleGluIleAsnSerSerProThr	335
2029	Db	GATTTCTTGTCGATTTCGGATTATGAAGTCAAGTTATTGTGAGATCAATGCTTCCAGAT	2088
336	Qy	MechisProSerThrProValThrAlaGlnLeuCysAlaGlnValGlnGluAspThr	355
2089	Db	TTCAAGCAAACTGGTAAAGACTTGAAGGACCTCATTTGATGATTTGTTGACGATCTGTC	2148
356	Qy	LysValAlaVal	359
2149	Db	AAGTATTGTGTT	2160

RESULT 7

```

US-10-932-182A-167
; Sequence 167, Application US/10932182A
; Publication No. US20060046253A1
; GENERAL INFORMATION:
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHIISA
; APPLICANT: KODAMA, YUKIO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: 030685-043
; CURRENT APPLICATION NUMBER: US/10/932,182A
; CURRENT FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 167
; LENGTH: 2250
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-10-932-182A-167

```

Alignment Scores:	
Pred. No.:	0.23
Pred. Score:	144.50
Length:	2250
Percent Similarity:	65
Best Local Similarity:	52
Mismatches:	124

Query Match:	5.1%	Indels:	83
DB:	9	Gaps:	10
US-10-635-977-2 (1-541) x US-10-932-182A-167 (1-2250)			
Qy	94	ThrGlnGlnTyrTyrSerLeuValHisGlyAspAlaPheIleSerAsnSer-----	110
Db	1264	ACTATCCAAACGATGATACGAAACAATCCAGATTCATTTTGAAGAAGCATATTAGAA	1323
Qy	111	-----ArgAsnTyrPheSerGlnCysGlnAlaLeuLeuAsnArgIleThr	125
Db	1324	TCGTTTACTATTGATTAGACTGCTGATTTTAGATGATGCATAGACGAAACTGG	1383
Qy	126	SerValAsnProGlnThrAspIleAspGlyLeuArgAsnIleTrpIleIleLysProAla	145
Db	1384	GAGTTAGTCAAGAAATTGGAAACGAAAGTAAAGGCAAAATGGTGGATTGTGAAACCAAGT	1443
Qy	146	AlaLysSerArgGlyArgAspIleValCysMetAspArgValGluGluIleLeuGluLeu	165
Db	1444	ATGAGCGATAAAGGTCAAGGTATTAGGGTATTCAAGACCATACAAGACCTACAAGCCATT	1503
Qy	166	-----	166
Db	1504	TTCGATTCTTTTGATGACGAAGATAGCGAAGCGAAGAGTGGAAATGACGACGACGCA	1563
Qy	167	AlaAlaAspHis-----ProLeuSerArgAspAsnLysTrpValVal	180
Db	1564	AATGCGAAATTCATGGATAATAACAAAGTAAATATCTCCAATTTGGCCACTTTATCAT	1623
Qy	181	GlnLysTrpIleGlnThrProLeuLeuIleCys-----AspThrLysPheAspIle	197
Db	1624	CAGAATTTCAAGAGCCCATTTAGTCTATCATCCATCGATGAATATAAGAAAGTTTCATATA	1683
Qy	198	ArgGlnTrpPheLeuValThrAspTrpAsnProLeuThrIleTrpPheTyrLysGluSer	217
Db	1684	AGGTGTTAT-----GTTGTTCTTAAAGGGGATTACAAAGTATTCGTATACGATAGGATG	1737
Qy	218	TyrLeuArgPheSerThrGlnArgPhe-----SerLeu	228
Db	1738	CTAGCACTATTTCCTGCTAAACCGTTTGTCCCTCTAGATCCTAACACATATTCGCTAACA	1797
Qy	229	AspLysLeuAspSerAlaIleHisLeuCysAsnAsnAlaValGlnLysTyrLeuLysAsn	248
Db	1798	GATTTCAAGAGATTGGAGTGCCATCTGACTATATACATGCTTGCAAAGC---AAGAAGAAA	1854
Qy	249	AspValGlyArgSerProLeuLeuProAlaHisAsnMetTrpThrSerThrArgPheGln	268
Db	1855	GATAAAGATTCTCTGTTTG-----GAGTTTGTAT	1884
Qy	269	GluTyrLeuGlnArgGlnGlyArgGlyAlaValTrpGlySerValIleTyrProSerMet	288
Db	1885	TCCTTAGAAGAA-----ATTCCAAACGAG	1908
Qy	289	LysLys-----AlaIleAlaHisAla	295
Db	1909	AAAAAGTTGAAATTAAAGAACAGATTCAATTGTCATACAAATGACGTTTTCTTAGCGGCA	1968
Qy	296	MetLysValAlaGlnAspHisValGluProArgLysAsnSerPheGluLeuTyrGlyAla	315
Db	1969	GTAATGTAACAGGCTAACTTCCAAACCGTTTACCAGTTCGTTTGGAGACCTATGTTGTA	2028
Qy	316	AspPheValLeuGlyArgAspPheArgProTrpLeuIleGluIleAsnSerSerProThr	335
Db	2029	GATTTCTTGTCGATTCGGATTATTGAAGTCAAGTATTTCGAGATCAATGCTTTCCAGAT	2088
Qy	336	MetHisProSerThrProValThrAlaGlnLeuCysAlaGlnValGlnGluAspThrIle	355
Db	2089	TTCAACCAACTGGTAAAGACTTGAAGGACCTCATTTGATGAATTTGTCAGCATCTGTC	2148
Qy	356	LysValAlaVal	359
Db	2149	AAGTATTGTGTT	2160

RESULT 8

US-10-750-185-41745
; Sequence 41745, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 41745
; LENGTH: 1319
; TYPE: DNA
; ORGANISM: Bovine 19866880883232
US-10-750-185-41745

Alignment Scores:
Pred. No.: 0.277 Length: 1319
Score: 141.00 Matches: 39
Percent Similarity: 49.2% Conservative: 23
Best Local Similarity: 31.0% Mismatches: 47
Query Match: 4.9% Indels: 17
DB: 8 Gaps: 5

US-10-635-977-2 (1-541) x US-10-750-185-41745 (1-1319)

```
QY 134 AspGlyLeuArgAsnIleTrpIleLeuLysProAlaAlaLysSerArgGlyArgAspIle 153
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 39 GAGGGAGAGGCAACGCTTGGATTGCAAGTCAATCAGCGGAGGCCAAAGGTGAAGGCATA 98
QY 154 ValCysMetAspArgValGluGluLeuLeuGluAlaAlaAspHisProLeuSer 173
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 99 CTCATCTTCTCAGAACTACAGAACTGTTGGATTTCATA----- 137
QY 174 ArgAspAsnLys-----TrpValValGlnLysTyrIleGluThrProLeuLeuIle 190
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 138 ---GACAAACCAAGCCAGATCGATCCAGAAATACCTGGAGCTCCCGTTGTTCCCTG 194
QY 191 -----CysAspThr-LysPheAspIleArgGlnTrpPheLeuValThrAspTrpAsnPr 208
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 195 GAAGCCTGGTCATCGCAAGTTTGACATCCGAGGCTGGGTCTTGGTGGAC-----TGTCA 248
QY 208 oLeuThrIleTrpPheTyrLysGluSerTyrLeuArgPheSerThrGlnArgPheSerLe 228
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 249 GTTTAATATCTACCTCTATAGTAGGGGTGCTCCAGACTGCTTCAGAACCGGTATCAGGT 308
QY 228 uAspLysLeu---AspSerAlaIleHisLeuCysAsnAsnAlaValGlnLysTyrLeuLy 247
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 309 GGATTATTTCCAAGATAAACCAACCACTTGACTAATCACTGCATTCAAAGGAGTACTC 368
QY 247 sAsnAspValGlyArg 252
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 369 AGAAGAACTATGGGAAG 384
```

RESULT 9

US-10-750-623-41745
; Sequence 41745, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David

; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 41745
; LENGTH: 1319
; TYPE: DNA
; ORGANISM: Bovine 19866880883232
US-10-750-623-41745

Alignment Scores:
Pred. No.: 0.277 Length: 1319
Score: 141.00 Matches: 39
Percent Similarity: 49.2% Conservative: 23
Best Local Similarity: 31.0% Mismatches: 47
Query Match: 4.9% Indels: 17
DB: 8 Gaps: 5

US-10-635-977-2 (1-541) x US-10-750-623-41745 (1-1319)

```
QY 134 AspGlyLeuArgAsnIleTrpIleLeuLysProAlaAlaLysSerArgGlyArgAspIle 153
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 39 GAGGGAGAGGCAACGCTTGGATTGCAAGTCAATCAGCGGAGGCCAAAGGTGAAGGCATA 98
QY 154 ValCysMetAspArgValGluGluLeuLeuGluAlaAlaAspHisProLeuSer 173
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 99 CTCATCTTCTCAGAACTACAGAACTGTTGGATTTCATA----- 137
QY 174 ArgAspAsnLys-----TrpValValGlnLysTyrIleGluThrProLeuLeuIle 190
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 138 ---GACAAACCAAGCCAGATCGATCCAGAAATACCTGGAGCTCCCGTTGTTCCCTG 194
QY 191 -----CysAspThr-LysPheAspIleArgGlnTrpPheLeuValThrAspTrpAsnPr 208
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 195 GAAGCCTGGTCATCGCAAGTTTGACATCCGAGGCTGGGTCTTGGTGGAC-----TGTCA 248
QY 208 oLeuThrIleTrpPheTyrLysGluSerTyrLeuArgPheSerThrGlnArgPheSerLe 228
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 249 GTTTAATATCTACCTCTATAGTAGGGGTGCTCCAGACTGCTTCAGAACCGGTATCAGGT 308
QY 228 uAspLysLeu---AspSerAlaIleHisLeuCysAsnAsnAlaValGlnLysTyrLeuLy 247
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 309 GGATTATTTCCAAGATAAACCAACCACTTGACTAATCACTGCATTCAAAGGAGTACTC 368
QY 247 sAsnAspValGlyArg 252
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 369 AGAAGAACTATGGGAAG 384
```

RESULT 10

US-11-124-368A-2877/c
; Sequence 2877, Application US/11124368A
; Publication No. US20050287559A1
; GENERAL INFORMATION:

; APPLICANT: Michele Cargill
; APPLICANT: James J. Devlin
; TITLE OF INVENTION: Genetic Polymorphisms Associated with
; TITLE OF INVENTION: Vascular Diseases, Methods of Detection and Uses Thereof
; FILE REFERENCE: CL001524
; CURRENT APPLICATION NUMBER: US/11/124,368A
; CURRENT FILING DATE: 2005-05-09
; PRIOR APPLICATION NUMBER: US 60/568,845
; PRIOR FILING DATE: 2004-05-07
; PRIOR APPLICATION NUMBER: US 60/625,936
; PRIOR FILING DATE: 2004-11-09
; NUMBER OF SEQ ID NOS: 21112

Qy	443	euGlyLeuLysGluGluLysGlyLeuProLeuAlaLeuLeuAlaProLeuArgGlyAlaA	463
Db	252	-----	222
		-----	-----
Qy	463	laGluSerGlyGlyAlaAlaGlnProThrArgThrLysAlaAlaGlyLysValGluLeuP	483
Db	221		
		-----	-----
		-----	-----
Qy	483	roAlaCysProCys-----ArgHisValAspSerGlnAlaProAsnThrGlyValPro	501
Db	174	--GCTGCTCCAGTGGTGGCGGTACAGCCACTGCTCGGCTCCGCTCGCAGCTGTCCAG	117
Qy	501	aAlaGlnProAlaLysSerTrpAspProAsnGlnLeuAsnAlaHisProLeuGluPro	521
Db	116	CTGTGGCTCTCTGCTGTGTGGACCCACACAGTCTCTGCCGGGCTCCCG-----	67
Qy	521	alLeuArgGlyLeuLysThrAlaGluGlyAlaLeuArgProProPro 536	
Db	66	-----	55

```

QY 410 SerAlaSerLeuLeuAspAlaGlnProLeuLysAlaArgGlyProSerAlaMetProAsp 429
Db 1112 TCCAGTCACTCTCAGCGAGCCAGCCT-----GGGCTCCAGGCTCCGCCACA 1065
QY 430 ProAlaGlnGlyProProSerProAlaLeuGlnArgAspLeuLysGlu-----447
Db 1064 TCCAGCGGGAGCCCGCCCGCCCGCGGCTCCGACACGCGGCTCCGGAGGCCCTT 1005
QY 448 ---GluLysGlyLeuProLeuAlaLeuLeuAlaProLeuArgGlyAlaAlaGluSerGly 466
Db 1004 GTGCGCGCGCGCTGCGCC---GGATGCGGGGCGCTCGCGGTCCACCGCGTCCCGGGGC 948
QY 467 -----GlyAlaAlaGln-----ProThrArgThrLys 475
Db 947 CGCGCGGGGGCTACCGAGCGGACTTCGGAGCGGCTCGACCTCCGCCCTCTCGTGCCTGC 888
QY 476 AlaAlaGlyLysValGluLeuProAlaCysProCysArgHisValAspSerGlnAlaPro 495
Db 887 GGGGAGGCTCAGCTCGGTGCGCCAGCTTCGCCAAGTCGTGCGACACCGCGCGCGCG 828
QY 496 -----AsnThrGlyValProValAlaGlnProAlaLysSerTrpAspProAsn 511
Db 827 CTGCTTCTCCCGATCAGGAGCACACAGCTTCGCCAAGTCGTGCGACACCGCGCGCG 768
QY 512 GlnLeuAsn-----AlaHisProLeuGluProValLeuArgGlyLeuLysThrAla 528
Db 767 CATTAGACTGAACATACGATACCTTTGGGCGCTTKDDDTGGTGGAGGGGCGGTC 708
QY 529 GluGly 530
Db 707 TCCGGT 702

RESULT 13
US-10-330-773-662
; Sequence 662, Application US/10330773
; Publication No. US20060040262A1
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; APPLICANT: Marc Malandro
; TITLE OF INVENTION: Novel Compositions and Methods in Cancer
; FILE REFERENCE: 529452001300
; CURRENT APPLICATION NUMBER: US/10/330,773
; CURRENT FILING DATE: 2002-12-27
; NUMBER OF SEQ ID NOS: 981
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 662
; LENGTH: 2754
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-330-773-662

Alignment Scores:
Pred. No.: 7.13 Length: 2754
Score: 126.50 Matches: 78
Percent Similarity: 29.4% Conservative: 15
Best local Similarity: 24.7% Mismatches: 110
Query Match: 4.4% Indels: 113
Db: 9 Gaps: 15

US-10-635-977-2 (1-541) x US-10-330-773-662 (1-2754)

QY 264 SerThrArgPheGlnGluTrpLeuGlnArgGlnGlyArgGlyAlaValTrpGlySerVal 283
Db 961 GCGACACGTTTGCACATCTCTA-----GGTTCAGCC 993
QY 284 IleTyrProSerMetLysLysAlaLeuAlaHisAlaMetLysValAlaGlnAspHisVal 303
Db 994 TTCACACCT-----GTTCTCCCATACGCTACAGTT-----1026
QY 304 GluProArgLysAsnSerPheGluLeuTyrGlyAlaAspPheValLeuGlyArgAspPhe 323
Db 1027 ---CCTCGTCTCT-----CTCAACAAACAACTCT 1050

```

```

QY 324 ArgProTrpLeuIleGluLeuAsnSerSerProThrMetHisProSerThr-----ProVal 342
Db 1051 CGACCT-----TCTTCTCTGTGTAACACACACACCTCTTCTCAGCTCCA 1092
QY 343 ThrAlaGlnLeuCysAlaGlnValGlnGluAspThrIleLysValAlaValAspArgSer 362
Db 1093 GTTGGAGAGTCTCTGTGCC-----1110
QY 363 CysAspIleGlyAsnPheGluLeuLeuTrpArgGlnPro-----ValValGluProPro 380
Db 1111 TGGCCTACTTCCAAATTTCTGCGCCCTCCCTCATCTCTCCATATATGATTAGCAGCC 1170
QY 381 ProPheSerGlySerAspLeuCysValAlaGlyValSerValArgAlaArgGln 400
Db 1171 CCTGGCAAGCTACT-----GSCCCACGCGCT 1197
QY 401 ValLeuProValCys-----405
Db 1198 GTCTCTTCCGGTTGTGTCTCTCTCTGTCGCCCAAAATGCTCCGTCACCAACAGCACCC 1257
QY 406 AsnLeuLysAlaSerAlaSerLeuLeuAspAlaGlnProLeuLysAlaArgGlyProSer 425
Db 1258 AATGGTCTGCTAGACTCTGTAAACATACCCAGGTCTTCCACCGCTACCTCAGGCGCAGCA 1317
QY 426 AlaMetProAspProAlaGlnGlyProProSerProAlaLeuGlnArgAspLeuGlyLeu 445
Db 1318 GCGCCACCTTCGCGCGCCACCGCCACCGCGCGCA-----1353
QY 446 LysGluGluLysGlyLeuProLeuAlaLeuLeuAlaProLeuArgGlyAlaAlaGluSer 465
Db 1354 ---CCACCACCGCGCTGCGCACCGCGCGCTGCTCCCTCGCTCCTCCTCCTCCTCCTCCT 1410
QY 466 GlyGlyAlaAlaGlnPro-----471
Db 1411 GATCAGAGCTTCTCTCTCTCCAGGACCCCTCTTGGCTCAACTCCCTCATCCCAAGCC 1470
QY 472 -----ThrArgThrLysAlaAlaGlyLysValGluLeuProAlaCysPro 486
Db 1471 AGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1530
QY 487 CysArgHisValAspSerGlnAlaProAsnThrGlyValProValAla---GlnProAla 505
Db 1531 ---GAGCTGGGAGACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1587
QY 506 LysSerTrpAspProAsnGlnLeu-----AsnAlaHisProLeuGluProVal 521
Db 1588 GAGTCGCCAACCCACAGGGCTTCTCTTGGGACCACTGCGCTCCGCGCCACACCCCTCTCTCT 1647
QY 522 LeuArgGlyLeuLysThrAlaGluGlyAlaLeuArgProProProGly 537
Db 1648 CTCCCATCAGGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1695

RESULT 14
US-11-136-527-1889
; Sequence 1889, Application US/11136527
; Publication No. US20050287570A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William M
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes
; FILE REFERENCE: 031896-041000 (AM101086)
; CURRENT APPLICATION NUMBER: US/11/136,527
; CURRENT FILING DATE: 2005-05-25
; PRIOR APPLICATION NUMBER: US 60/574,294
; PRIOR FILING DATE: 2005-05-26
; NUMBER OF SEQ ID NOS: 362830
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1889
; LENGTH: 5354
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-11-136-527-1889

```



```
Alignment Scores:
Pred. No.: 14.8 Length: 5354
Score: 126.00 Matches: 82
Percent Similarity: 36.7% Conservative: 21
Best Local Similarity: 29.2% Mismatches: 90
Query Match: 4.4% Indels: 88
DB: 14 Gaps: 20

US-10-635-977-2 (1-541) x US-11-136-527-1889 (1-5354)
QY 306 ArgLysAsnSerPheGluLeuTyrglyAlaAspPheValLeuGlyArgAspPheArgPro 325
DB 4424 AGGAGGAATCTTCAGAGTTCAGAGGGGAC-----4456
QY 326 TrpLeuIleGluLeuAsnSerProThrMetHisProSerThrProValThr-----343
DB 4457 -----CAAGCAACATCCCTTCACCTCCCTCGCCCTGCTT 4492
QY 344 ---AlaGlnLeuCysAlaGlnValGlnGluAspThrIleLysValAlaValAspArgSer 362
DB 4493 CCCTCCAGCTCTCTCGCGGGCC-----ACCGTTCCAACTGCGTCTTCTTC 4543
QY 363 CysAspIleGlyAsnPheGluLeu-----TrpArgGlnProValValGluProPro 380
DB 4544 TGGGAGCAAGACATTTCAGATGTTGTACGGTGGCGGCGACTCAGGCAGCGCCCTT 4603
QY 381 Pro-----PheSerGlySerAspLeuValAlaGlyValSerValArgArg 396
DB 4604 CCCCTAACGGCTCTTCTCCGCTCC-----ATTGCTCCAGTTCCCTCCCTGCG-- 4651
QY 397 AlaArgGlnValLeuPro-ValCysAsnLeuLysAlaSerAlaSerLeuLeuAspAl 416
DB 4652 -----CCTCCCTCTCTCT-----CCACCTCGGGAGCGCTGGGGGT 4690
QY 416 aGlnPro-----LeuLysAlaArgGlyPro-----SerAlaMe 427
DB 4691 GTGGCCAGGGGCGCGGTATAAGTCCCGGAGCGGTCCCGGAGCGGCTCAGCCYCT 4750
QY 427 tProAspProAlaGlnGlyProProSerProAlaLeuGlnArgAspLeuGlyLeuLysGl 447
DB 4751 GCCCCCGCCGCTGCTGGCGCGCGGAGGATGACAGCGAGCGC---CTCGGTGTC---CA 4804
QY 447 uGluLysGlyLeuProLeuAlaLeuAlaProLeu-----ArgGl 461
DB 4805 GGCTCGCTGCTCAGCGTCTTGCTAACTTCCCTTGGCTCCGCTCCAGCTGCGGGG 4864
QY 461 yAlaAlaGluSerGly-----466
DB 4865 CCGCACCGTGTCTCTCTCGCTTCGGGTCTGGTCTTAAGCTCTCTGGGAGTGCCTCC 4924
QY 467 -GlyAlaAlaGlnProThrArgThrLysAlaAlaGlyLysValGluLeuProAlaCys-- 485
DB 4925 GGGTTCCTTGACGCGCGCTCGAGCGCGG-----CTCCCTTCTCGC 4969
QY 486 ---ProCys-ArgHisValAspSerGlnAlaProAsnThrGlyValProValAlaGlnP 504
DB 4970 CCCACCTCTGTTCTCACCATCTCTCCCTCGGCTCGT---GGCGCCTCTGCTCGCC 5026
QY 504 roAlaLysSerTrpAspProAsnGlnLeuAsnAlaHisProLeuGluProValLeuArg 524
DB 5027 GCAGGGCTCGCGAGCTCGGGCA---CGAAGCATTCACCTGCCCTCTTGTCTCCGAG 5083
QY 524 lyLeuLysThrAlaGluGlyAlaLeuArgProProProGly-----GlyLysGlySer 541
DB 5084 GA-----GAAGCTGGCGCGTGGCGCCCCCGGCTGTGTGAGGAGCTGGTGG 5133
```

RESULT 15

```
US-10-995-561-13381/c
; Sequence 13381, Application US/10995561
; Publication No. US2005027054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
```

```
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: C001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13381
; LENGTH: 14248
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-13381
```

```
Alignment Scores:
Pred. No.: 38.5 Length: 14248
Score: 126.00 Matches: 49
Percent Similarity: 34.1% Conservative: 13
Best Local Similarity: 26.9% Mismatches: 64
Query Match: 4.4% Indels: 56
DB: 8 Gaps: 9
```

US-10-635-977-2 (1-541) x US-10-995-561-13381 (1-14248)

```
QY 375 ProValValGluProProPheSerGlySer-----AspLeuCysVal 389
DB 6063 CTTCCCGAGCGCCCGCGCTTTCGGATCCCCCTCCAGCTCTCGTCTCTGTCTC 6004
QY 390 AlaGlyValSerValArgAlaArgGlnValLeuProValCysAsnLeuLysAla 409
DB 6003 CTTGGGATC-----5995
QY 410 SerAlaSerLeuLeuAspAlaGlnProLeuLysAlaArgGlyProSerAlaMetProAsp 429
DB 5994 -----CTGGTTCCTTGGAGTCACACCCACC 5968
QY 430 ProAlaGlnGlyProProSerProAlaLeuGlnArgAsp---LeuGlyLeuLysGluGlu 448
DB 5967 CCGGCTCAA---CCGGGCTCCCGGACCTCGCGGCGGACCCGGCGGCTGACTCCGAGT 5911
QY 449 LysGlyLeuProLeuAlaLeuLeuAlaProLeuArgGlyAlaAlaGluSerGlyAla 468
DB 5910 CGCGGCTCCCGGTCTCTGCGCGCGCGCGGCTCCGCTCCGCGCGGGTCCC 5851
QY 469 AlaGlnProThrArgThrLysAlaAlaGlyLysValGluLeuProAlaCysProCysArg 488
DB 5850 GGCCCGCGCGCGCGCGCGC-----TGACCTGTCTCC 5818
QY 489 HisValAspSerGlnAlaProAsnThrGlyValProValAlaGlnPro---AlaLysSer 507
DB 5817 TGCTCTCACGCCACCGCGCTCGCTCCCGCCCCCTCCCGCGCGCGCGGCTCGCGGC 5758
QY 508 TrpAspProAsnGlnLeuAsnAlaHisProLeu-----GluProValLeu--- 522
DB 5757 TGGCGGTCTCCGCTCTGTCTGTCCGTCCGTGGGTCCGCGAGACCGACCTGGGCTAGGC 5698
QY 523 ArgGlyLeuLysThrAlaGluGlyAlaLeu-----ArgProProPro 536
DB 5697 AGGGGTCTAGACCGCGGAGGGGGCGGCACGAGGAGGAACCGGGCCCGGGGACCGCCA 5638
QY 537 GlyGly 538
DB 5637 GGGGA 5632
```

Search completed: April 4, 2006, 12:29:52
Job time : 4850 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 4, 2006, 09:54:42 ; Search time 1192 Seconds

(without alignments)
3753.132 Million cell updates/sec

Title: US-10-635-977-2

Perfect score: 2854

Sequence: 1 MASSILKVVVSHQSCRSR.....LRGLTARGALRPPPGKGKS 541

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_p2n.model -DEV=xlh
-Q=/abss/ABSSWEB_spool/US10635977/runat_04042006_103543_10259/app_query.fasta_1
-DB=Published Applications NA Main -QFMT=fastap -SUFFIX=p2n.rnpbm
-MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1
-MATRIX=blosum62 -TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR SCORE=pct
-THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext
-HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abss07
-USER=US10635977 @CEN 1.1.1026 @runat_04042006_103543_10259 -NCPU=6 -ICPU=3
-NO MMAP -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications NA Main:

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2854	100.0	1838	7	US-10-615-659-1
2	2854	100.0	1838	7	US-10-615-977-1
3	2843	99.6	3554	7	US-10-615-659-12
4	2843	99.6	3554	7	US-10-635-977-12
5	2698.5	94.6	1859	7	US-10-615-659-10
6	2698.5	94.6	1859	7	US-10-635-977-10
7	2693	94.4	3465	7	US-10-615-659-11

8	2693	94.4	3465	7	US-10-635-977-11	Sequence 11, Appl	
9	2689	94.2	1939	7	US-10-615-659-9	Sequence 9, Appl	
10	2689	94.2	1939	7	US-10-635-977-9	Sequence 9, Appl	
c	11	1802	63.1	101270	8	US-10-723-860-631	Sequence 631, App
12	1299	45.5	726	7	US-10-615-659-3	Sequence 3, Appl	
13	1299	45.5	726	7	US-10-635-977-3	Sequence 3, Appl	
14	1070.5	37.5	2380	7	US-10-275-595A-39	Sequence 39, Appl	
15	1020	35.7	2538	5	US-10-102-524-1765	Sequence 1765, Ap	
16	1020	35.7	2553	9	US-10-956-157-2016	Sequence 2016, Ap	
17	1020	35.7	2553	9	US-10-756-149-1971	Sequence 1971, Ap	
18	1020	35.7	3001	5	US-10-037-270-494	Sequence 494, App	
19	1020	35.7	3001	6	US-10-117-722-494	Sequence 494, App	
20	1020	35.7	3001	9	US-10-122-851-494	Sequence 494, App	
21	998	35.0	2611	3	US-09-925-298-199	Sequence 199, App	
22	998	35.0	2611	3	US-10-102-806-199	Sequence 199, App	
23	983.5	34.5	2848	6	US-10-108-260A-1766	Sequence 1766, Ap	
24	926	32.4	1673	9	US-10-956-157-4121	Sequence 4121, Ap	
25	912.5	32.0	1400	9	US-10-956-157-9356	Sequence 9356, Ap	
26	785	27.5	1400	9	US-10-956-157-7251	Sequence 7251, Ap	
c	27	760	26.6	490	6	US-10-029-386-24894	Sequence 24894, A
28	733	25.7	2979	10	US-11-097-143-23726	Sequence 23726, A	
c	29	678	23.8	526	5	US-10-027-632-135928	Sequence 135928, A
c	30	678	23.8	526	6	US-10-027-632-135928	Sequence 135928, A
c	31	675	23.7	526	5	US-10-027-632-135927	Sequence 135927, A
c	32	675	23.7	526	6	US-10-027-632-135927	Sequence 135927, A
c	33	556.5	19.5	5728	10	US-11-097-143-23725	Sequence 23725, A
34	536.5	18.8	2543	10	US-11-097-143-23414	Sequence 23414, A	
35	514.5	18.0	4615	10	US-11-097-143-23413	Sequence 23413, A	
36	478.5	16.8	1958	9	US-10-956-157-4903	Sequence 4903, Ap	
37	409.5	14.3	436	7	US-10-242-535A-7660	Sequence 7660, Ap	
38	409.5	14.3	436	7	US-10-085-783A-7660	Sequence 7660, Ap	
39	405	14.2	418	3	US-09-983-965-2032	Sequence 2032, Ap	
40	394	13.8	1728	9	US-10-450-763-5962	Sequence 5962, Ap	
41	390.5	13.7	2250	9	US-10-450-763-5187	Sequence 5187, Ap	
42	364.5	12.8	755	5	US-10-027-632-135929	Sequence 135929, A	
43	364.5	12.8	755	5	US-10-027-632-135930	Sequence 135930, A	
44	364.5	12.8	755	6	US-10-027-632-135929	Sequence 135929, A	
45	364.5	12.8	755	6	US-10-027-632-135930	Sequence 135930, A	

ALIGNMENTS

RESULT 1

US-10-615-659-1

; Sequence 1, Application US/10615659
; Publication No. US20040157234A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; CURRENT FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; NUMBER OF SEQ IDS NOS: 102
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 1838
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (153)..(1775)
US-10-615-659-1

Alignment Scores:
Pred. No.: 9.2e-287 Length: 1838
Score: 2854.00 Matches: 541
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 7 Gaps: 0


```
Query Match: 100.0% Indels: 0
DB: 7 Gaps: 0
US-10-635-977-2 (1-541) x US-10-635-977-1 (1-1838)

QY 1 MetAlaSerSerIleLeuLysTrpValValSerHisGlnSerCysSerArgSerSerArg 20
DB 153 ATGGCATCTCAGCATCTCAAGTGGGTGGTGCACCCAGAGCTGCAGCAGGAGCAGCAGA 212
QY 21 SerLysProArgAspGlnArgGluGluAlaGlySerSerAspLeuSerSerArgGlnAsp 40
DB 213 AGCAAGCCAGGAGCAGAGGAGGAGGCGCGAGCAGCGACTTGAGCAGCAGGCAAGAT 272
QY 41 AlaGluAsnAlaGluAlaLysLeuArgGlyLeuProGlyGlnLeuValAspIleAlaCys 60
DB 273 GCTGAAATGCTGAGGCAAGCTCAGGGGCCCTCCCGGGGCAGCTTGTGGACATCGCTGC 332
QY 61 LysValCysGlnAlaTyrLeuGlyGlnLeuGluHisGluAspIleAspThrSerAlaAsp 80
DB 333 AAGGTGTGCCAGCCCTACTCGGGCAGCTGGAGCATGAGACATCGACACGTACAGAT 392
QY 81 AlaValGluAspLeuThrGluAlaGluTrpGluAspLeuThrGlnGlnTyrTrpSerLeu 100
DB 393 GCGGTGGAGGACTCACTGAGGCGGAGTGGGAGGACCTGACCCAGCAGTACTACTCCCTC 452
QY 101 ValHisGlyAspAlaPheIleSerAsnSerArgAsnTyrPheSerGlnCysGlnAlaLeu 120
DB 453 GTTCATGGCGATGCTTTTCACTCCAAATTCAGAAATTTACTTTTCAGTGGCCAGGCTCTG 512
QY 121 LeuAsnArgIleThrSerValAsnProGlnThrAspIleAspGlyLeuArgAsnIleTrp 140
DB 513 CTGAATAGATACACGTCTGTGAACCTCAGACGGACATTGACGGGCTCCGGAAATCTGG 572
QY 141 IleIleLysProAlaAlaLysSerArgGlyArgAspIleValCysMetAspArgValGlu 160
DB 573 ATTATAAGCCCGCGCCCAAGTCCCGGGCCGAGACATAGTGTGACATGACCGTGGAG 632
QY 161 GluIleLeuGluLeuAlaAlaAspHisProLeuSerArgAspAsnLysTrpValVal 180
DB 633 GAGATCTTGGAGCTGGAGCTGCAGACCACTCTTTCCAGGACCAACAAGTGGGTGGTC 692
QY 181 GlnLysTyrIleGluThrProLeuLeuIleCysAspThrLysPheAspIleArgGlnTrp 200
DB 693 CAGAAATGATCATGACGCCCGCTGCTCATCTGTGACACCAAGTTTCGACATCAGACAGTGG 752
QY 201 PheLeuValThrAspTrpAsnProLeuThrIleTrpPheTyrLysGluSerTyrLeuArg 220
DB 753 TTCTCTCGTCACGGACTGGAAACCCCTGACCATCTGGTTCTACAGGAGAGTTACTTGGG 812
QY 221 PheSerThrGlnArgPheSerLeuAspLysLeuAspSerAlaIleHisLeuCysAsnAsn 240
DB 813 TTCTCACTCAGCGCTTCTCCCTGGACCAAGCTGGACAGCGCCATCCACTGTGCAACAAC 872
QY 241 AlaValGlnLysTyrLeuLysAsnAspValGlyArgSerProLeuLeuProAlaHisAsn 260
DB 873 GCCGTCCAGAGTACTCTGAAGAATGATGGGCCGACGCCCTGCTGCCCGGCACACAAAC 932
QY 261 MetTrpThrSerThrArgPheGlnGluTyrLeuGlnArgGlnGlyArgGlyAlaValTrp 280
DB 933 ATGTGGACCAACACAGGTTCCAGGAGTACCTGACGCGCCAGGGCCGTGGCGCGTGGG 992
QY 281 GlySerValIleTyrProSerMetLysLysAlaIleAlaHisAlaMetLysValAlaGln 300
DB 993 GGCAGCGTCATCTACCCCTCCATGAAGAGGCCATCGCCACCGCATGAAGTGGCCCGAG 1052
QY 301 AspHisValGluProArgLysAsnSerPheGluLeuTyrGlyAlaAspPheValLeuGly 320
DB 1053 GACACGTGGAGCCTCGCAAGAACACTTTGAGTCTACGGGGCTGACTTCGCTCTTGGG 1112
QY 321 ArgAspPheArgProTrpLeuIleGluIleAsnSerSerProThrMetHisProSerThr 340
DB 1113 AGGACCTTCAGGCCCTGCTGATCGAGATCAATTCAGCCCAACATGACCCCGTCCAGC 1172
```

341 ProValThrAlaGlnLeuCysAlaGlnValGlnGlnAspThrIleLysValAlaValAsp 360
1173 CCGGTACGGCCAGCTGTGTGCACAGGTGCAGGAGGACACCATCAAGGTGCGCGTGGAC 1232
361 ArgSerCysAspIleGlyAsnPheGluLeuLeuTrpArgGlnProValValGluProPro 380
1233 CGCAGCTGTGACATCGGCAACTTCGAGCTCTCTGTGTGAGCAGCCCGGTGTTGAGCGCGCC 1292
381 ProPheSerGlySerAspLeuCysValAlaGlyValSerValArgAlaArgArgGln 400
1293 CCATTTCAGCGGTCCGACCTCTGCGTGGGGCGCTCAGTGTGAGGAGAGCCAGAGGCGAG 1352
401 ValLeuProValCysAsnLeuLysAlaSerAlaSerLeuLeuAspAlaGlnProLeuLys 420
1353 GTGCTGCCCGCTGTCAACCTCAAGGCCCTCGGCTCGCTGTGTGACGCGCAGCGCTGAAG 1412
421 AlaArgGlyProSerAlaMetProAspProAlaGlnGlnGlyProProSerProAlaLeuGln 440
1413 GCACGGGGCCCTCGGCCATTCGACACCTGTCGCCAGGACCCCATCACAGCTCTCCAG 1472
441 ArgAspLeuGlyLeuLysGluLysGlyLeuProLeuAlaLeuAlaProLeuArg 460
1473 CGGACCTTGGGACTGAAGAGAGAGAGGGGCTCCCTCGCTGCTGTCGACACCTTAAG 1532
461 GlyAlaAlaGluSerGlyGlyAlaAlaGlnProThrArgThrLysAlaAlaGlyLysVal 480
1533 GGGGAGCCGAGAGCGGTGGAGCCGACAGCCGACCCGACCCAAAGCTGCTGGGAGGTG 1592
481 GluLeuProAlaCysProCysArgHisValAspSerGlnAlaProAsnThrGlyValPro 500
1593 GAGCTCCCGGCTGCTGCCCTGTGCCACGTGACAGTACGGGCCCAACACCGGTGTCCCC 1652
501 ValAlaGlnProAlaLysSerTrpAspProAsnGlnLeuAsnAlaHisProLeuGluPro 520
1653 GTAGCCAGCCCGCCAAAGCTGGATCCAAACAGCTAAATGCGCACCCGCTGGAGCT 1712
521 ValLeuArgGlyLeuLysThrAlaGluAlaLeuArgProProGlyGlyLysGly 540
1713 GTGCTCGGGGCTTGAAGACAGCAGAGGGCGCTGCTGCGCCCGCGAGGAAAGGT 1772
541 Ser 541
1773 TCA 1775

RESULT 3
US-10-615-659-12
; Sequence 12, Application US/10615659
; Publication No. US20040157234A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; TITLE OF INVENTION: TYROSINE-LIGASE-LIKE PROTEIN, BGS42
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; CURRENT FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 12
; LENGTH: 3554
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-615-659-12

Alignment Scores:
Pred. No.: 3,24e-285 Length: 3554
Score: 2843.00 Matches: 539
Percent Similarity: 99.6% Conservative: 0
Best Local Similarity: 99.6% Mismatches: 2
Query Match: 99.6% Indels: 0
DB: 7 Gaps: 0


```
Db 47 ATGGCATCCAGCATCCTCAAGTGGTGGTCAGCCACGAGCTGCAGCAGGACGACGAGA 106
Qy 21 SerLysProArgAspGlnArgGluAlaGlySerSerAspLeuSerSerArgGlnAsp 40
Db 107 AGCAAGCCAGGACACAGAGGAGGAGCGCGGAGCAGCGACCTGAGCAGCAGCGCAAGAT 166
Qy 41 AlaGluAsnAlaGluAlaLysLeuArgGlyLeuProGlyGlnLeuValAspLeuAlaCys 60
Db 167 GCTGAAATAGCTCAGGCAAGCTCAGGGGGCTCCCGGGGAGCTTGTGGACATCGCGGTGC 226
Qy 61 LysValCysGlnAlaTyrLeuGlyGlnLeuCluHisGluAspLeuAspThrSerAlaAsp 80
Db 227 AAGGTGTGCAGAGCTTACCTGGGGCAGCTGGAGCATGAGGACATCGACAGCTCAGCAGAT 286
Qy 81 AlaValGluAspLeuThrGluAlaGluTyrGluAspLeuThrGlnGlnTyrTyrSerLeu 100
Db 287 GCCGTGGAGACCTCACTGAGGCGGAGTGGAGGACCTGACCCAGCAGTACTACTCCCTC 346
Qy 101 ValHisGlyAspAlaPheIleSerAsnSerArgAsnTyrPheSerGlnCysGlnAlaLeu 120
Db 347 GTTCA-----TGCCAGGCTCTG 363
Qy 121 LeuAsnArgIleThrSerValAsnProGlnThrAspIleAspGlyLeuArgAsnIleTyr 140
Db 364 CTGATAGATACAGTCTGTGAACCTCAGACGGACATTTACCGGGCTCCGGAAACATCTGG 423
Qy 141 IleIleLysProAlaLysSerArgGlyArgAspIleValCysMetAspArgValGlu 160
Db 424 ATTATAAGCCCGCGCCAAAGTCCCGGGGCGGAGACATAGTGTGCATGGACCGTGTGGAG 483
Qy 161 GluIleLeuGluLeuAlaAlaAspHisProLeuSerArgAspAsnIleTyrValVal 180
Db 484 GAGATCTGGAGCTGGCAGCTGCAGACCAACCTCTTTCCAGGGACAAACAAGTGGGTGTC 543
Qy 181 GlnLysTyrIleGluThrProLeuLeuIleCysAspThrLysPheAspIleArgGlnTyr 200
Db 544 CAGAAGTACATCGAGACCGCGTGTCTCATCTGTGACACCAAGTTTCGACATCAGCAGTGG 603
Qy 201 PheLeuValThrAspTyrAsnProLeuThrIleTyrPheTyrLysGluSerTyrLeuArg 220
Db 604 TTCTCTGTACGAGCTGAACCCCTGACCATCTGGTTCTACAGGAGAGTACTTTCGCG 663
Qy 221 PheSerThrGlnArgPheSerLeuAspLysLeuAspSerAlaIleHisLeuCysAsnAsn 240
Db 664 TTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGACAGCGCCATCCACCTGTGCAACAC 723
Qy 241 AlaValGlnLysTyrLeuLysAsnAspValGlyArgSerProLeuLeuProAlaHisAsn 260
Db 724 GCCGTCAGAAAGTACCTGAAGATGATGTGGCGCGCAGCCCCCTGCTGCCCGCACACAAC 783
Qy 261 MetTyrThrSerThrArgPheGlnGlnTyrLeuGlnArgGlnGlyArgGlyAlaValTyr 280
Db 784 ATGTGGACCAAGCAGCAGGTTCCAGAGTACCTGCAGCGCCAGGGCGCGTGGCGTGTGG 843
Qy 281 GlySerValIleTyrProSerMetLysLysAlaIleAlaHisAlaMetLysValAlaGln 300
Db 844 GGCAGCGCTCATCTACCGCTCATGAAGAAGCCATCGCCACGCAATGAAGGTGGCCAG 903
Qy 301 AspHisValGluProArgLysAsnSerPheGluLeuTyrGlyAlaAspPheValLeuGly 320
Db 904 GACCACGTGGAGCCTCGCAAGAACAGCTTTGAGGCTTACGGGGCTGACTTCTGCTTGGG 963
Qy 321 ArgAspPheArgProTyrLeuIleGluIleAsnSerSerProThrMetHisProSerThr 340
Db 964 AGGAGTTCAGGCCCTGGCTGATCGAGATCAATTCAGGCCCAACCATGCACCCGCTCCACG 1023
Qy 341 ProValThrAlaGlnLeuCysAlaGlnValGlnGluAspThrIleLysValAlaValAsp 360
Db 1024 CCGGTCAAGCCCAAGCTGTGTGCACAGTGCAGAGGACACCAATCAAGGTGGCGGTGGAC 1083
Qy 361 ArgSerCysAspIleGlyAsnPheGluLeuLeuTyrArgGlnProValValGluProPro 380
Db 1084 CGCAGCTGTGACATCGGCNACTTCAGACTCTCTGTGGAG-----CGGTGGTTGACCCGCC 1139
```

```
Qy 381 ProPheSerGlySerAspLeuCysValAlaAlaGlyValSerValArgArgAlaArgArgGln 400
Db 1140 CCATTTCAGCGGGTCCAGACTCTGCGTGGCGGGCGTCAAGTGTGAGGAGAGCAGGAGCAG 1199
Qy 401 ValLeuProValCysAsnLeuLysAlaSerAlaSerLeuLeuAspAlaGlnProLeuLys 420
Db 1200 GTGCTGCCCGTCTGCAACCTCAAGGCTCGGGCTCGCTGTGGACGCGCAGCGCTGAAG 1259
Qy 421 AlaArgGlyProSerAlaMetProAspProAlaGlnGlyProProSerProAlaLeuGln 440
Db 1260 GCACGGGGCCCTCGGCCATGCGCAGACCCCTGCGCAGGAGCCGCCATCACCAGCTCTCCAG 1319
Qy 441 ArgAspLeuGlyLeuLysGlnGlyLysGlyLeuProLeuAlaLeuLeuAlaProLeuArg 460
Db 1320 CGGGACTTGGGACTGAAGGAAGAGAGGGGCTCCCTCGCTTGTGTGGACCCCTTAAGG 1379
Qy 461 GlyAlaAlaGluSerGlyGlyAlaAlaGlnProThrArgThrLysAlaAlaGlyLysVal 480
Db 1380 GGGGACCGCGAGAGCGGTGGACCGCACAGCCCGCACCAAGCTGCTGGGAAGGTG 1439
Qy 481 GluLeuProAlaCysProCysArgHisValAspSerGlnAlaProAsnThrGlyValPro 500
Db 1440 GAGCTCCCGCGCTGCGCTGTGCGCACGTGGACAGTCAGGCCCAACACCGGTGTCCCC 1499
Qy 501 ValAlaGlnProAlaLysSerTyrAspProAsnGlnLeuAsnAlaHisProLeuGluPro 520
Db 1500 GTAGCCCGCGCCCAAGAGTGGGATCCAAACCACTAANTGAGCACCCCGCTGGAGCCT 1559
Qy 521 ValLeuArgGlyLeuLysThrAlaGluGlyAlaLeuArgProProGlyGlyLysGly 540
Db 1560 GTGCTCGGAGCTGAAGACAGAGGGCGCGCTGCTCGCGCCGCGAGGAAAGGT 1619
Qy 541 Ser 541
Db 1620 TCA 1622
```

RESULT 6

```
US-10-635-977-10
; Sequence 10, Application US/10635977
; Publication No. US20040171131A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE OF INVENTION: TYROSINE-LIGASE-LIKE PROTEIN, BGS42
; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635,977
; PRIOR FILING DATE: 2003-08-07
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S.10/615,659
; PRIOR FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 10
; LENGTH: 1859
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-635-977-10
```

```
Alignment Scores:
Pred. No.: 1,52e-270 Length: 1859
Score: 2698.50 Matches: 522
Percent Similarity: 96.5% Conservative: 0
Best Local Similarity: 96.5% Mismatches: 4
Query Match: 94.6% Indels: 17
DB: 2 Gaps: 2
```

US-10-635-977-2 (1-541) x US-10-635-977-10 (1-1859)

```
Qy 1 MetAlaSerSerIleLeuLysTyrValValSerHisGlnSerCysSerArgSerArg 20
Db 47 ATGCATCCAGCATCCTCAAGTGGTGGTTCAGCCACGAGCTGCAGGAGGAGCAGAGA 106
```



```
Db 1738 AGCAAGCCCGGACGACAGGAGGAGGCGCGGAGCGACGACCTGAGCAGCAGGCAAGAT 1797
Qy 41 AlaGluAsnAlaGluAlaLysLeuArgGlyLeuProGlyGlnLeuValAspIleAlaCys 60
Db 1798 GCTGAAATGCTGAGGCAAGCTCAGGGGCGCTCCCGGGGCGAGCTTGTGGACATCGCGTGC 1857
Qy 61 LysValCysGlnAlaTyrLeuGlyGlnLeuGluHisGluAspIleAspThrSerAlaAsp 80
Db 1858 AAGGTGTGCGAGCGCTTACCTGGGGCAGCTGGAGCATGAGGACATCGACACGTCAGCAGAT 1917
Qy 81 AlaValGluAspLeuThrGluAlaGluTyrGluAspLeuThrGlnGlnTyrTyrSerLeu 100
Db 1918 GCGGTGGAGGACTTCACTGAGGCGGAGTGGAGGACCTGACCCAGCAGTACTACTCCCTC 1977
Qy 101 ValHisGlyAspAlaPheIleSerAsnSerArgAsnTyrPheSerGlnCysGlnAlaLeu 120
Db 1978 GTTCATGGCGATGCTTTCATCTCCAATTCAGAAATTACTTTTCGAGTGCCAGGCTCTG 2037
Qy 121 LeuAsnArgIleThrSerValAsnProGlnThrAspIleAspGlyLeuArgAsnIleTyr 140
Db 2038 CTGAATAGAATCAGCTCTGTGAACCTCAGACGGAATTCAGCGGCTCCGGAACATCTGG 2097
Qy 141 IleIleLysProAlaLysSerArgGlyArgAspIleValCysMetAspArgValGlu 160
Db 2098 ATTATTAAGCCCGCGCCAGTCCCGGGGCGG----- 2129
Qy 161 GluIleLeuGluLeuAlaAlaAspHisProLeuSerArgAspAsnLysTyrValVal 180
Db 2130 -----AGGACAAACAAGTGGTGGTC 2150
Qy 181 GlnLysTyrIleGluThrProLeuLeuIleCysAspThrIlyspPheAspIleArgGlnTyr 200
Db 2151 CAGAAGTACATCAGACGCGCGCTGCTCATCTGTGACACCAAGTTCGACATTCAGACAGTGG 2210
Qy 201 PheLeuValThrAspTyrAsnProLeuThrIleTyrPheTyrLysGluSerTyrLeuArg 220
Db 2211 TTCCTCGTCAGGACTGGAAACCCCTGACCACTGTTCTTACAAAGAGATTACTTTCGGG 2270
Qy 221 PheSerThrGlnArgPheSerLeuAspLysLeuAspSerAlaIleHisLeuCysAsnAsn 240
Db 2271 TTCTCAACTCAGCGCTTCTCCCTGGCAAGCTGACAGCGCCATCCACCTGTGCCAACAC 2330
Qy 241 AlaValGlnLysTyrLeuLysAsnAspValGlyArgSerProLeuLeuProAlaHisAsn 260
Db 2331 GCGGTCCAGAAATGATGTGGCGCGCAGCGCCCTGCTGCGCGCGCACACAC 2390
Qy 261 MetTyrThrSerThrArgPheGlnGluTyrLeuGlnArgGlnGlyArgGlyAlaValTyr 280
Db 2391 ATGTGGACCAAGCAGGTTCCAGGAGTACCTGCGAGCGCCAGGCGCGTGGCGCGTGG 2450
Qy 281 GlySerValIleTyrProSerMetLysLysAlaIleAlaHisAlaMetLysValAlaGln 300
Db 2451 GGCAGCGCTCATCCCGCTCATGAAGAAGGCCATCGCCACGCCATGAAGGTGGCCAG 2510
Qy 301 AspHisValGluProArgLysAsnSerPheGluLeuTyrGlyAlaAspPheValLeuGly 320
Db 2511 GACCACGTGAGCGCTCGCAAGAACAGCTTTGAGCTCTACGCGGCTGACTTCGCTTGGG 2570
Qy 321 ArgAspPheArgProTyrLeuIleGluIleAsnSerSerProThrMetHisProSerThr 340
Db 2571 AGGAGCTTCAGGCGCTCGCATCGAGATCAATTCAGGCCCCACCATGCAACCGCTCCAGC 2630
Qy 341 ProValThrAlaGlnLeuCysAlaGlnValGlnGluAspThrIleLysValAlaValAsp 360
Db 2631 CCGGTTCAGCGCCAGCTGTGTGCACAGGTGCAGGAGGACCATCAAGGTGGCGGTGGAC 2690
Qy 361 ArgSerCysAspIleGlyAsnPheGluLeuLeuTyrArgGlnProValValGluProPro 380
Db 2691 CGCAGCTGTGACATCGGCAACTTCGAGCTCTGTGTGAGGCGCGGTTGGTGTAGCGCGCC 2750
Qy 381 ProPheSerGlySerAspLeuCysValAlaGlyValSerValArgAlaArgArgGln 400
Db 2751 CCATTACGCGGCTCCGACCTCTGCTGCGGCGCGTCAAGTGTGAGGAGGACCGAGGCGAG 2810
```

```
Qy 401 ValLeuProValCysAsnLeuLysAlaSerAlaSerLeuLeuAspAlaGlnProLeuLys 420
Db 2811 GTGCTGCCGCTCTGCAACCTCAAGGCTCGGCTCGCTTTGGACGCGGCGCTGAAG 2870
Qy 421 AlaArgGlyProSerAlaMetProAspProAlaGlnGlyProProSerProAlaLeuGln 440
Db 2871 GCACGGGGCCCTCGGCCATGCCAGACCTGCCCAGGACCCCATCAACAGCTCTCCAG 2930
Qy 441 ArgAspLeuGlyLeuLysGluLysGlyLeuProLeuAlaLeuLeuAlaProLeuArg 460
Db 2931 CGGCACTTGGGACTCAAGGAGAGAGGGGCTCCCTTGGCTTGTGTCACCTTAAGG 2990
Qy 461 GlyAlaAlaGluSerGlyGlyAlaAlaGlnProThrArgThrLysAlaAlaGlyLysVal 480
Db 2991 GGGGAGCGCGAGGCGGTGGAGCCGACAGCCACCCGACCAAGCTCTGTTGGAGGTG 3050
Qy 481 GluLeuProAlaCysProCysArgHisValAspSerGlnAlaProAsnThrGlyValPro 500
Db 3051 GAGCTCCCGGCTGCCCCCTGTGCCACGTTGGACAGTCAGGCCCAACACCGGTGTCCCC 3110
Qy 501 ValAlaGlnProAlaLysSerTrpAspProAsnGlnLeuAsnAlaHisProLeuGluPro 520
Db 3111 GTAGCCAGCCCGCCAAAAGCTGGGATCCAAAACCAAGCTAAATGAGCACCCGCTGGAGCCT 3170
Qy 521 ValLeuArgGlyLeuLysThrAlaGluGlyAlaLeuArgProProGlyGlyLysGly 540
Db 3171 GTGCTCGGAGGCTGAAGACAGCAGAGGGGCGCTGCGTCCGCGCCCGAGGAAAAGGT 3230
Qy 541 Ser 541
Db 3231 TCA 3233
```

RESULT 8

```
US-10-635-977-11
; Sequence 11, Application US/10635977
; Publication No. US20040171131A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; TITLE OF INVENTION: TYROSINE-LIGASE-LIKE PROTEIN, BG842
; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635,977
; PRIOR FILING DATE: 2003-08-07
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S. 10/615,659
; PRIOR FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 11
; LENGTH: 3465
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-635-977-11
```

```
Alignment Scores:
Pred. No.: 1,36e-269 Length: 3465
Score: 2693.00 Matches: 516
Percent Similarity: 95.4% Conservative: 0
Best Local Similarity: 95.4% Mismatches: 3
Query Match: 94.4% Indels: 23
DB: 7 Gaps: 1
```

US-10-635-977-2 (1-541) x US-10-635-977-11 (1-3465)

```
Qy 1 MetAlaSerSerIleLeuLysTyrValValSerHisGlnSerCysSerArgSerArg 20
Db 1678 ATGCATCCAGCATCTCTCAAGTGGTGTGAGCCACAGAGCTGCAGAGGAGCAGAGA 1737
Qy 21 SerLysProArgAspGlnArgGluAlaGlySerSerAspLeuSerSerArgGlnAsp 40
Db 1738 AGCAAGCCCGAGGACCAAGAGGAGGAGGCGCGGAGCAGACCTTGAGCAGGAGCAAGAT 1797
```

QY 41 AlaGluAsnAlaGluAlaLysLeuArgGlyLeuProGlyGlnLeuValAspIleAlaCys 60
DB 1798 GCTGAAATATGCTCAGGCAAGCTCAGGGGCTCCCGGGCAGCTTGTGGACATCGCGTGC 1857
QY 61 LysValCysGlnAlaTyrLeuGlyGlnLeuGluHisGluAspIleAspThrSerAlaAsp 80
DB 1858 AAGGTGTGCCAGCCCTACCTTGGGGCAGCTGGAGCATGAGACATCCACACGTCAGCAGAT 1917
QY 81 AlaValGluAspLeuThrGluAlaGluTyrGluAspLeuThrGlnGlnTyrIleValSerLeu 100
DB 1918 GCGTGGAGGACCTCACTGAGGCCGAGTGGAGGACCTGACCCAGCAGTACTACTCCCTC 1977
QY 101 ValHisGlyAspAlaPheIleSerAsnSerArgAsnTyrPheSerGlnCysGlnAlaLeu 120
DB 1978 GTTCATGGGGATGCTCTTCAATCTCAATTCAGAAATTTACTTTTCGCAGTGCACGCTCTG 2037
QY 121 LeuAsnArgIleThrSerValAsnProGlnThrAspIleAspClyLeuArgAsnIleTyr 140
DB 2038 CTGAATAGAAATCACGTCTGTGAACCTCTCAGACGACATTCAGCGGCTCCGGAAACATCTGG 2097
QY 141 IleIleLysProAlaAlaLysSerArgGlyArgAspIleValCysMetAspArgValGlu 160
DB 2098 ATTATTAAGCCCGGGCCAGTCCCGGGCGG----- 2129
QY 161 GluIleLeuGluLeuAlaAlaAspHisProLeuSerArgAspAsnLysTyrValVal 180
DB 2130 -----AGGACAAACAAGTGGTGGTC 2150
QY 181 GlnLysTyrIleGluThrProLeuLeuIleCysAspThrLysPheAspIleArgGlnTyr 200
DB 2151 CAGAAATGATCATCGAGACCGCGTGTCTCATCTGTGACACCAAGTTCGACATCAGACAGTGG 2210
QY 201 PheLeuValThrAspTyrAsnProLeuThrIleTyrPheTyrLysGluSerTyrLeuArg 220
DB 2211 TTCCTCGTACGAGCTGGAAACCCCTGACCATCTGGTCTTCAAGGAGAGTACTTGGCG 2270
QY 221 PheSerThrGlnArgPheSerLeuAspLysLeuAspSerAlaIleHisLeuCysAsnAsn 240
DB 2271 TTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGACAGCGCCATCCACCTGTGTCAACAAC 2330
QY 241 AlaValGlnLysTyrLeuLysAsnAspValGlyArgSerProLeuLeuProAlaHisAsn 260
DB 2331 GCGGTCCAGAACTACTGAAGATGATGTGGCCGCGCAGCGCCCTGTGTCGCCGACACAAAC 2390
QY 261 MetTyrThrSerThrArgPheGlnGluTyrLeuGlnArgGlnGlyArgGlyAlaValTyr 280
DB 2391 ATGTGACACGACCAAGTTCCAGGAGTACCTTCAGCGCCAGCGCGCTGGCGCGCTGTGG 2450
QY 281 GlySerValIleTyrProSerMetLysLysAlaIleAlaHisAlaMetLysValAlaGln 300
DB 2451 GGCAGCGTCATCTACCCGTCATGAAGAAGGCCATCGCCACGCCATGAAGTGGCCCGAG 2510
QY 301 AspHisValGluProArgLysAsnSerPheGluLeuTyrGlyAlaAspPheValLeuGly 320
DB 2511 GACCACTGGAGCCCTCGCAAGAACAGCTTGTAGCTCTACGGGCTGACTTCGTCCTTGGG 2570
QY 321 ArgAspPheArgProTyrLeuIleGluIleAsnSerSerProThrMetHisProSerThr 340
DB 2571 AGGGAATTCAGGCGCTTGGTGTGAGATCAATTCAGGCCCCACCATGCACCCGCTCCAG 2630
QY 341 ProValThrAlaGlnLeuCysAlaGlnValGlnGluAspThrIleLysValAlaValAsp 360
DB 2631 CCGGTACGGCCCGAGCTGTGTGCACAGGTGCAGGAGGACCATCAAGGTGGCCGTGGAC 2690
QY 361 ArgSerCysAspIleGlyAsnPheGluLeuLeuTyrArgGlnProValValGluProPro 380
DB 2691 CGCAGCTGTGATCATCGCAACTTCGAGCTCTCTGTGGAGCGCAGCGGTGTGGCGCGCC 2750
QY 381 ProPheSerGlySerAspLeuCysValAlaGlyValSerValArgArgAlaArgGln 400
DB 2751 CCATTCAGCGGTCTCCACCTCTGCGTGGCGGGCGTGTGAGGAGGCCAGGAGGCGAG 2810

QY 401 ValLeuProValCysAsnLeuLysAlaSerAlaSerLeuLeuAspAlaGlnProLeuLys 420
DB 2811 GTGCTGCCGCTCTGCAACCTCAAGGCTCGGCCCTCGCTGTTGGACGCGCAGCGCTGAAG 2870
QY 421 AlaArgGlyProSerAlaMetProAspProAlaGlnGlyProProSerProAlaLeuGln 440
DB 2871 GCACGGGGCCCTCGGCCATGCGACACCTGCCAGGACCCCATCACAGCTCTCCAG 2930
QY 441 ArgAspLeuGlyLeuLysGluGlyLysGlyLeuProLeuAlaLeuLeuAlaProLeuArg 460
DB 2931 CGGCACTTGGGACTGAGGAGAGAGGGGCTCCCTCGCTTGTGGCACCTTAAGG 2990
QY 461 GlyAlaAlaGluSerGlyGlyAlaAlaGlnProThrArgThrLysAlaAlaGlyLysVal 480
DB 2991 GGGGACGCGAGAGCGGTGGAGCGCACAGCCACCGCAAAAGCTGCTGGGAAGGTG 3050
QY 481 GluLeuProAlaCysProCysArgHisValAspSerGlnAlaProAsnThrGlyValPro 500
DB 3051 GAGCTCCCGGCTCGCCCTGTGCGCACGTGCGAGTCAAGCCCAACACCGGTGTCCCC 3110
QY 501 ValAlaGlnProAlaLysSerTyrAspProAsnGlnLeuAsnAlaHisProLeuGluPro 520
DB 3111 GTAGCCACCGCCGCAAAAGCTGGGATCCAAACACAGCTAAATGAGCACCCGCTGGAGCCT 3170
QY 521 ValLeuArgGlyLeuLysThrAlaGluGlyAlaLeuArgProProGlyGlyLysGly 540
DB 3171 GTGCTCGGAGCTTGAAGACAGACAGAGGGCGCGCTGCGTCCGCGCCGCGAGGAGGAGT 3230
QY 541 Ser 541
DB 3231 TCA 3233

RESULT 9

US-10-615-659-9
; Sequence 9, Application US/10615659
; Publication No. US20040157234A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; TITLE OF INVENTION: TYROSINE-LIGASE-LIKE PROTEIN, BGS42
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; CURRENT FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 9
; LENGTH: 1939
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-615-659-9

Alignment Scores:
Pred. No.: 1586-269 Length: 1939
Score: 2689.00 Matches: 515
Percent Similarity: 95.2% Conservatives: 0
Best Local Similarity: 95.2% Mismatches: 4
Query Match: 94.2% Indels: 23
DB: 7 Gaps: 1

US-10-635-977-2 (1-541) x US-10-615-659-9 (1-1939)

QY 1 MetaLysSerIleLeuLysTyrValValSerHisGlnSerCysSerArgSerArg 20
DB 153 ATGGCATCCAGCATCTCAAGTGGTGTTCAGCCACCCAGAGCTGCAGCAGGAGCAGAGA 212
QY 21 SerLysProArgAspGlnArgGluAlaGlySerSerAspLeuSerArgGlnAsp 40
DB 213 AGCAAGCCAGGGACCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 272
QY 41 AlaGluAsnAlaGluAlaLysLeuArgGlyLeuProGlyGlnLeuValAspIleAlaCys 60

```

Db      273  GCTGAAATGCTGAGGCAAGCTCAGGGGCTCCGGGGCAGCTTGTGGACATCGCGTGC 332
Qy      61   LysValCysGlnAlaTyrLeuGlyGlnLeuGluHisGluAspIleAspThrSerAlaAsp 80
Db      333  AAGGTGTGCGCAGGCTACCTGGGGCAGCTGGAGCATGAGACATCGACAGCTCAGCAGAT 392
Qy      81   AlaValGluAspLeuThrGluAlaGluTrpGluAspLeuThrGlnGlnTyrTyrSerLeu 100
Db      393  GCCGTGGAGACCTCACTGAGGCCGAGTGGGAGACCTGACCCAGCAGTAGTACTACTCCCTC 452
Qy     101   ValHisGlyAspAlaPheIleSerAsnSerArgAsnTyrPheSerGlnCysGlnAlaLeu 120
Db     453  GTTCATGGCGATGCTTTCATCTCAATTCAAGAAATTACTTTTCGAGTGCAGGCTCTCG 512
Qy     121   LeuAsnArgIleThrSerValAsnProGlnThrAspIleAspGlyLeuArgAsnIleTrp 140
Db     513  CTGAATAGAAATCACGTCTGTGAACCTCTACAGCGAATTGACCGGCTCCGGAACATCTGG 572
Qy     141   IleIleIysProAlaAlaIysSerArgGlyArgAspIleValCysMetAspArgValGlu 160
Db     573  ATTATAAGCCCGCGCAAGTCCCGGGCGC-----AGGCACAACAGTGGTGTGC 604
Qy     161   GluIleLeuGluAlaAlaAspHisProLeuSerArgAspAsnIysTrpValVal 180
Db     605  -----AGGCACAACAGTGGTGTGC 625
Qy     181   GlnIysTyrIleGluThrProLeuLeuIleCysAspThrIysPheAspIleArgGlnTrp 200
Db     626  CAGAAGTACATCGAGACCGCGTGCATCTGTGACACCAAGTTCGACATCAGACAGTGG 685
Qy     201   PheLeuValThrAspTrpAsnProLeuThrIleTrpPheTyrIysGluSerTyrLeuArg 220
Db     686  TTCCTCGTCAGGACCGGCAACCCCTGACCATCTGGTTCACAGGAGAGATTACTTGC GG 745
Qy     221   PheSerThrGlnArgPheSerLeuAspLysLeuAspSerAlaIleHisLeuCysAsnAsn 240
Db     746  TTTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGACAGCGCCATCCACCTGTGCAACAC 805
Qy     241   AlaValGlnIysTyrLeuLysAsnAspValGlyArgSerProLeuLeuProAlaHisAsn 260
Db     806  ACCGTCAGCAAGTACCTGAAGATGATGTGGCGCGCAGCCCTGCTGCTGCCGCACACAC 865
Qy     261   MetTrpThrSerThrArgPheGlnIuTyrLeuGlnArgGlnGlyAlaValTrp 280
Db     866  ATGTGGACACGACCGAGTTCCAGGAGTACCTCGAGCGCCAGGGCGCGTGGCGGTGG 925
Qy     281   GlySerValIleTyrProSerMetLysLysAlaIleAlaHisAlaMetLysValAlaGln 300
Db     926  GGCAGCGTCACTTACCCGTCATGAGAGAGGCCATCGCCACGCCATGAAGTGGCCCGAG 985
Qy     301   AspHisValGluProArgLysAsnSerPheGluLeuTyrGlyAlaAspPheValLeuGly 320
Db     986  GACCACGTGGAGCCTCGCAAGAACAGCTTTGAGCTCTACGGGGTGACTTCGTCCTTGGG 1045
Qy     321   ArgAspPheArgProTrpLeuIleGluIleAsnSerSerProThrMetHisProSerThr 340
Db    1046  AGGAGACTTCAGGGCCTGGCTGATCGAGATCAATTTCCAGCCGCCACCATGCAACCCGTCACG 1105
Qy     341   ProValThrAlaGlnLeuCysAlaGlnValGlnGluAspThrIleLysValAlaValAsp 360
Db    1106  CCGGTACGGCCCAAGCTGTGTGCACAGTGCAGAGAGGACACCATCAAGTGGCGCTGGAC 1165
Qy     361   ArgSerCysAspIleGlyAsnPheGluLeuLeuTrpArgGlnProValValGluProPro 380
Db    1166  CGCAGCTGTGACATCGGCAACTTCGAGCTCTGTGGAGGCGAGCGGTGGTTGAGCCGCC 1225
Qy     381   ProPheSerGlySerAspLeuCysValAlaGlyValSerValArgAlaArgArgGln 400
Db    1226  CCATTACGGGGTCCGACCTCTGGGTGGCGGGCGTCAAGTGTGAGAGAGAGCCAGAGGCAG 1285
Qy     401   ValLeuProValCysAsnLeuLysAlaSerAlaSerLeuLeuAspAlaGlnProLeuLys 420
Db    1286  GTGCTGCCCGTCTGCACCTCAAGGCTTCGGCCTTCGCTGTGTGGACGGCGCAGCCGCTGAG 1345

; Sequence 9, Application US/10635977
; Publication No. US20040171131A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; TITLE OF INVENTION: TYROSINE-LIGASE-LIKE PROTEIN, BG942
; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635,977
; CURRENT FILING DATE: 2003-08-07
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S.10/615,659
; PRIOR FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 9
; LENGTH: 1939
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-635-977-9

Alignment Scores:
Pred. No.: 1,58e-269 Length: 1939
Score: 2689.00 Matches: 515
Percent Similarity: 95.2% Conservative: 0
Best Local Similarity: 95.2% Mismatches: 4
Query Match: 94.2% Indels: 23
DB: 7 Gaps: 1

US-10-635-977-2 (1-541) x US-10-635-977-9 (1-1939)
Qy      1   MetAlaSerSerIleLeuLysTrpValValSerHisGlnSerCysSerArgSerArg 20
Db     153  ATGCGATCCAGCATCCTCAAGTGGTGGTCCAGCCACCAGAGCTCCAGCAGAGCAGCAGA 212
Qy     21   SerIysProArgAspGlnArgGluAlaGlySerSerAspLeuSerSerArgGlnAsp 40
Db     213  AGCAAGCCCGAGGAGACAGAGAGGCGCGGAGACGCGACCTGAGCAGCAGCAGCAGAT 272
Qy     41   AlaGluAsnAlaGluAlaLysLeuArgGlyLeuProGlyGlnLeuValAspIleAlaCys 60
Db     273  GCTGAAATGCTGAGGCAAGCTCAGGGGCTCCGGGGCAGCTTGTGGACATCGCGTGC 332
```

```
QY 61 LysValCysGlnAlaTyrLeuGlyGlnLeuGluHisGluAspIleAspThrSerAlaAsp 80
Db 333 AAGGTGTGCAGCGCTTACCTGGGGCAGCTGGAGCATGAGACATCCACACGTCCAGCAT 392
QY 81 AlaValGluAspLeuThrGluAlaGluTyrGluAspLeuThrGlnGlnTyrTyrSerLeu 100
Db 393 GCCGTGGAGACCTTACTGAGCGCCGAGTGGAGGACCTGACCCAGCAGTACTTCCCTC 452
QY 101 ValHisGlyAspAlaPheIleSerAsnSerArgAsnTyrPheSerGlnCysGlnAlaLeu 120
Db 453 GTTCATGGCGATGCTTTCATCTCCATTCAAGAAATTAATTTTCGAGTGGCCAGGCTCTG 512
QY 121 LeuAsnArgIleThrSerValAsnProGlnThrAspIleAspGlyLeuArgAsnIleTyr 140
Db 513 CTGAATAGAAATCACGCTGTGTAAACCTTCAGACGGACATTCAGCGGCTCCGGAAATCTCG 572
QY 141 IleIleLysProAlaAlaLysSerArgGlyArgAspIleValCysMetAspArgValGlu 160
Db 573 ATTATTAAGCCCGCGCCGANGTCCCGGGCCG----- 604
QY 161 GluIleLeuGluLeuAlaAlaAspHisProLeuSerArgAspAsnLysTyrValVal 180
Db 605 -----AGGACACACAGTGGGTGTC 625
QY 181 GlnLysTyrIleGluThrProLeuLeuIleCysAspThrLysPheAspIleArgGlnTyr 200
Db 626 CAGAAAGTATCATCGAGCGCGCTGCTCATCTGTGACACCAAGTTCGACATCAGACAGTGG 685
QY 201 PheLeuValThrAspTyrAsnProLeuThrIleTyrPheTyrLysGluSerTyrLeuArg 220
Db 686 TTCCTCGTCAGCGACTGAACCCCTGACCATCTGGTTCTACAGGAGAGTTACTTGGGG 745
QY 221 PheSerThrGlnArgPheSerLeuAspLysLeuAspSerAlaIleHisLeuCysAsnAsn 240
Db 746 TTCTCACTCAGCGCTTTCCTCTGACAGCTGGACGGCCCATCCACTGTGCAACAC 805
QY 241 AlaValGlnLysTyrLeuLysAsnAspValGlyArgSerProLeuLeuProAlaHisAsn 260
Db 806 ACCGTCCAGAAAGTACCTGAAGAATGATGTGGCCCGCAGCCCTCTCTCCCGCACAC 865
QY 261 MetTyrThrSerThrArgPheGlnGlnTyrLeuGlnArgGlnGlyArgGlyAlaValTyr 280
Db 866 ATGTGGACAGACACAGGTTCCAGGAGTACTCTGACGCCAGCGCCGCGTGGCGCGCTGG 925
QY 281 GlySerValIleTyrProSerMetLysLysAlaIleAlaHisAlaMetLysValAlaGln 300
Db 926 GGCAGCGTCATCTACCCGTCCATGAGAGGCGCATCGCCACGCCATGAAGTGGCCCGAG 985
QY 301 AspHisValGluProArgLysAsnSerPheGluLeuTyrGlyAlaAspPheValLeuGly 320
Db 986 GACCACGTGGAGCTTCGCAAGAACAGCTTTGAGCTCTACGGGGCTGACTTCGTCTTGGG 1045
QY 321 ArgAspPheArgProTyrPheLeuIleGluLeuAsnSerProThrMetHisProSerThr 340
Db 1046 AGGGAATTCAGGCGCTGCTGATCGAGATCAATTCAGCCCGCACCATCCAGCCGCTCCAG 1105
QY 341 ProValThrAlaGlnLeuCysAlaGlnValGlnAspThrIleLysValAlaValAsp 360
Db 1106 CCGTTCAGCGCCCGAGCTGTGTGACAGGTGAGGAGGACACCATCAGGTGGCCGTGGAC 1165
QY 361 ArgSerCysAspIleGlyAsnPheGluLeuLeuTyrArgGlnProValValGluProPro 380
Db 1166 CGCAGCTGTGACATCGGCAACTTCGAGCTCTGTGGAGCGAGCGGTGTGGAGCGGCC 1225
QY 381 ProPheSerGlySerAspLeuCysValAlaGlyValSerValArgAlaArgGln 400
Db 1226 CCATTCAGCGGCTCGCACTCTGCGTGGCGGCGTCACTGTGAGGAGACCCAGGAGGCGAG 1285
QY 401 ValLeuProValCysAsnLeuLysAlaSerAlaSerLeuLeuAspAlaGlnProLeuLys 420
Db 1286 GTGCTGCCCGCTGTCAACCTCAAGGCGCTCGCGCTCGCTGTGTGGAGCGCGCGAGCTGAAG 1345
```

```
QY 421 AlaArgGlyProSerAlaMetProAspProAlaGlnGlyProProSerProAlaLeuGln 440
Db 1346 GCACGGGGCCCCCTCGCCCATGCGACCCCTGCCAGGAGACCCCATCACCAGCTCTCCAG 1405
QY 441 ArgAspLeuGlyLeuLysGluGluLysGlyLeuProLeuAlaLeuLeuAlaProLeuArg 460
Db 1406 CCGGAGCTTGGGACTGAAGGAAGAGAGGGGCTCCCGCTTGGTGGCACCCTTAAGG 1465
QY 461 GlyAlaAlaGluSerGlyGlyAlaAlaGlnProThrArgThrLysAlaAlaGlyLysVal 480
Db 1466 GGGGCGAGCGAGAGCGGTGGAGCCGACAGCCACCCAGAGCTGCTGGGAAGGTG 1525
QY 481 GluLeuProAlaCysProCysArgHisValAspSerGlnAlaProAsnThrGlyValPro 500
Db 1526 GAGCTCCCGGCGCTGCCCTGTGCCACGCTGGACAGTCAAGCCCAACACCCGCTGCCCC 1585
QY 501 ValAlaGlnProAlaLysSerTyrAspProAsnGlnLeuAsnAlaHisProLeuGluPro 520
Db 1586 GTAGCCCAAGCCCGCCCAAGCTGGGATCCAAACACAGCTAAATGAGCACCCGCTGGAGCCT 1645
QY 521 ValLeuArgGlyLeuLysThrAlaGluGlyAlaLeuArgProProGlyGlyLysGly 540
Db 1646 GTGCTGGAGCGCTGAAGACAGCAGAGGGCGGCTGCGTCCGCCCGCGGAGGAAAGGT 1705
QY 541 Ser 541
Db 1706 TCA 1708
```

RESULT 11
US-10-723-860-631/c
; Sequence 631, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Nataasha
; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; FILE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators
; FILE REFERENCE: 05882.0193.NFUS01
; CURRENT APPLICATION NUMBER: US/10/723,860
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429,739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 631
; LENGTH: 101270
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-723-860-631

Alignment Scores:
Pred. No.: 1.18e-174 Length: 101270
Score: 1802.00 Matches: 438
Percent Similarity: 32.2% Conservative: 1
Best Local Similarity: 32.1% Mismatches: 3
Query Match: 63.1% Indels: 923
DB: 8 Gaps: 3

US-10-635-977-2 (1-541) x US-10-723-860-631 (1-101270)

```
QY 100 LeuValHisGlyAspAlaPheIleSerAsnSerArgAsnTyrPheSerGlnCysGlnAla 119
Db 47222 CTTTGTAGTGGGATGCTTTCATCTCCAAATTCAGAAATTAATTTTCGAGTGGCAGGCT 47163
QY 120 LeuLeuAsnArgIleThrSerValAsnProGlnThrAspIleAspGlyLeuArgAsnIle 139
Db 47162 CTGCTGAATAGAAATCACTGTCGAAACCTCAGACGGACATTGACGGGCTCCGGAACATC 47103
QY 140 TrpIleIleLysProAlaAlaLysSerArgGlyArg----- 151
Db 47102 TGGATTATAAAGCCCGCGCGCAAGTCCGCGGGCCG-AGGTGAGTCCCTGCTGCGGATGC 47044
```

Qy	151	-----	151	
Db	47043	TCCCTGCA	CCGCGTT	CATAGCCAA
Qy	151	-----	151	-----
Db	46983	GGACAGTTT	CAGGCCCA	CACACCA
Qy	151	-----	151	-----
Db	46923	GGCGTGGGGT	CAGGACT	CACGGCGCT
Qy	151	-----	151	-----
Db	46863	TCCGATCA	CAGGGCAGGAT	CCGGGTT
Qy	151	-----	151	-----
Db	46803	CAAACTGGT	CTATGCTT	TCCTTTT
Qy	151	-----	151	-----
Db	46743	TGAACGGAGT	TTCACTT	CTTGTG
Qy	151	-----	151	-----
Db	46683	TGCAACCT	CCACCTT	CCAGGTT
Qy	151	-----	151	-----
Db	46623	ACTACAGG	CGCCACCA	CACCCGGT
Qy	151	-----	151	-----
Db	46563	CCATGTTG	CGCAGGCT	GGTCTT
Qy	151	-----	151	-----
Db	46503	CCAAAGTG	CTGGGATT	TACAGGCAT
Qy	151	-----	151	-----
Db	46443	AGTGTCT	GTGTCCTG	CCCTTCA
Qy	151	-----	151	-----
Db	46383	GTTCTGTT	GGGGAGAG	CTGGGGG
Qy	151	-----	151	-----
Db	46323	TAAGCCTT	CCCTGCT	CCCCAT
Qy	151	-----	151	-----
Db	46263	AGCCAGTG	AGGTCTC	AGGGCCA
Qy	152	-----	152	-----
Db	46203	CCGCTGCA	AGTGCT	GTGCATG
Qy	175	MetAsp	ArgVal	GlulGlul
Db	46143	ATGGACCT	GTGGAGG	ATCTGG
Qy	195	AsnLys	TrpVal	GlnLys
Db	46083	AACAAGT	GGTGGT	CCAGAA
Qy	215	AspIle	ArgGln	TrpPhe
Db	46023	GACATCA	GACAGT	GGTTC
Qy	232	GluSer	TyrLeu	ArgPhe


```
Db 44883 GACCGCCCTTGGTGCCTCCTGGGCCAAGGGAGGCTGAGCACCTGCCCTCGTCCCCC 44824
QY 233 ---SerAlaIleHisLeuCysAsnAenAlaValGlnLysTyLeuLysAsnAepValGly 251
Db 44823 TGCAGCGCCATCCACCTGTGCAACACACGCCGTCCAGAACTACCTGAAGAATGATGTGGC 44764
QY 252 ArgSerProLeuLeuProAlaHisAsnMetTTPThrSerThrArgPheGlnGluTyLeu 271
Db 44763 CCGAGCCCTCCTGTGCCCCGACACAACTGTGGACCAAGCAGGTTCCAGAGTACCTG 44704
QY 272 GlnArgGlnGlyArgGlyAlaValTTPGlySerValIleTyProSerMetLysLysAla 291
Db 44703 CAGCGCCAGGCGCGTGGCCCTGTGGGCGAGGTCACTACCCGTCCATGAAGAGGCC 44644
QY 292 IleAlaHisAlaMetLysValAlaGlnAepHisValGluProArgLysAsnSerPheGlu 311
Db 44643 ATCGGCCACGCCATGAAGGTGGCCCCAGGACCAACGTCGAGCCTCGCAAGAACAGCTTTGAG 44584
QY 312 LeuTyGlyAlaAaspPheValLeuGlyArgAaspPheArgProTTPLeuIleGluIleAsn 331
Db 44583 CTCTACGGGGCTGACTTCGTCTTGGAGGGAGCTTTCAGGCCCTTGGCTGATCGAGATCAAT 44524
QY 332 SerSerProThrMetHisProSerThrProValThrAlaGlnLeuCysAlaGlnValGln 351
Db 44523 TCCAGGCCACCATGCACCCGTCACGCCGCTCAAGGCCAGCTGTGTGCACAGGTGCAG 44464
QY 352 GluAaspThrIleLysValAlaValAaspArgSerCysAaspIleGlyAsnPheGluLeu 371
Db 44463 GAGGACACCATCAAGGTGGCCGTGACCGCAGCTGTGACATCGGCAACTTCGAGCTCCTG 44404
QY 372 TTPArg----- 373
Db 44403 TGGAGGAGGTGAGCCACGCCCGCCCTGGGGACTTTGGTGGCCGCCGCCCATGTGGATG 44344
QY 373 ----- 373
Db 44343 CTTTGGCACCTGGTCACTGGGCGAGGTGGAGCTGCAGGCCGCCAGTCAGCGGGGTCTA 44284
QY 373 ----- 373
Db 44283 TGGAGGACACCATCTCTGTGACCATGCTCGTCCCGTGTAGGGGCCCGAGGCCCTGCC 44224
QY 373 ----- 373
Db 44223 CTGCTCTCCATGGGGCTCAGTGCCTCGTCCGACCGAGGAAGGGGCTACTCCAGGCGTG 44164
QY 373 ----- 373
Db 44163 CGGGCGAGCGCCCGCTCACTCAGGCTCAGTGCCTCTGTCCAGCTTTCTCCTCTTTG 44104
QY 373 ----- 373
Db 44103 TTCCAGGCTCACCCAGTCTGACTGTGTGTAGAGTGGGGCTCCATGCCCGCAGCCTC 44044
QY 373 ----- 373
Db 44043 CAGGCCACCTATGCTCTCTTTGGCTTGCAGTGCAGTGCCTCTCCACTCTGAGGATCT 43984
QY 373 ----- 373
Db 43983 TTGGGGGCTCCTGTGCTGGGATGTGTGAGTCAITGGGAAGGCTGGCCAAATGCC 43924
QY 373 ----- 373
Db 43923 TCTCTTTGGGCTCTCTGTGGCTTAACCTCGGGAAGAGCAGCTGAGTCCCTGGGTGT 43864
QY 373 ----- 373
Db 43863 GGGGCCCCCAGGCTGACCCCAAGCCTCAGGAAGCCTTGGTTCGCCCGAGGTTCTCCAG 43804
QY 373 ----- 373
Db 43803 AGCCTGGGGGCTCCTGTGGGGGTCTGGTGGGACCCAGGTGTTCACCTGTTCACTCTCTGAG 43744
```

```
QY 373 ----- 373
Db 43743 TCAGGTGGGGGTCTGTGCCCCGACCCAGCCCCCTCTCTGCACAGACCCCTCTCTGCAGGAGCCC 43684
QY 374 ----- 374
Db 43683 CCCCTATCCCGCCACAGCCACCCCAACCTGTCTCTCCACACAGCCCGTGGTTGA 43624
QY 378 uProProPheSerGlySerAaspLeuCysValAlaGlyValSerValArgAlaArg 398
Db 43623 GCCGCCCCCATTCACGGGGTCCAGCTCTGCGTGGCGGCGTCAGTGTGAGGAGCCAG 43564
QY 398 GArgGlnValLeuProValCysAsnLeuLysAlaSerAlaSerLeuLeuAaspAlaGlnPr 418
Db 43563 GAGGAGGTGTCTGCCGTCTGCAACCTCAAGGCCCTCGGCTCGCTGTTGGACGCGCAGCC 43504
QY 418 oleuLysAlaArgGlyProSerAlaMetProAaspProAlaGlnGlyProProSerProAl 438
Db 43503 GCTGAAGGCACGGGGCCCTCGGCCATGCCAGACCTGCCAGGAGCCCCCATCACCAGC 43444
QY 438 aLeuGlnArgAaspLeuGlyLeuLysGluGlyLysGlyLeuProLeuAlaLeuAlaPr 458
Db 43443 TCTCAGCGGAGCTTGGGACTGAAGGAAGAGAGGGGCTCCCTCTGGCCTTGTGTCACC 43384
QY 458 oleuArgGlyAlaAlaGluSerGlyGlyAlaAlaGlnProThrArgThrLysAlaAlaG 478
Db 43383 CTTAAGGGGGCGAGCCGAGAGCGGTGGAGCGCCACAGCCACCCCAAGCTGCTGG 43324
QY 478 yLysValGluLeuProAlaCysProCysArgHisValAaspSerGlnAlaProAaspThrGl 498
Db 43323 GAAAGTGGAGCTCCCGGCTGCTCCCTGTGCGCACGTGAGACAGTCAGGGCCCCAACACCGG 43264
QY 498 yValProValAlaGlnProAlaLysSerTTPAaspProAenGlnLeuAenAlaHisProle 518
Db 43263 TGTCCCGTAGCCAGCCCGCCAAAAGCTGGGATCCAAACAGCTAAATGCCACCCGCT 43204
QY 518 uGluProValLeuArgGlyLeuLysThrAlaGluGlyAlaLeuArgProProGlyGly 538
Db 43203 GGAGCTGTGTCTGCGGGCTGTAAGACAGACAGAGCGCGCTGCGCTCCGCCCGCCGAGG 43144
QY 538 yLysGlySer 541
Db 43143 AAAAGGTAAC 43134
```

RESULT 12
US-10-615-659-3
; Sequence 3, Application US/10615659
; Publication No. US20040157234A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE OF INVENTION: TYROSINE-LIGASE-LIKE PROTEIN, BGS42
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; PRIOR FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3
; LENGTH: 726
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(726)
US-10-615-659-3

Alignment Scores:
Pred. No.: 3,31e-125 Length: 726
Score: 1299.00 Matches: 242
Percent Similarity: 100.0% Conservative: 0

Best Local Similarity: 100.0%		Mismatches: 0
Query Match: 45.5%		Indels: 0
DB: 7		Gaps: 0
US-10-635-977-2 (1-541) x US-10-615-659-3 (1-726)		
QY	133 IleAspGlyLeuArgAsnIleTrpIleIleLysProAlaAlaLysSerArgGlyArgAsp 152	
DB	1 ATTGACGGGCTCGGAACATCTGGATTATAAGCCCGCGCAAGTCCCGGGCCGAGAC 60	
QY	153 IleValCysMetAspArgValGluGluIleLeuGluLeuAlaAlaAspHisProLeu 172	
DB	61 ATAGTGTGATGACCGTGTGGAGGATCCTGGAGCTGGACATCGACACCACTCTTT 120	
QY	173 SerArgAspAsnLysTrpValValGlnLysTyrIleGluThrProLeuLeuIleCysAsp 192	
DB	181 ACCAAGTTCGACATCAGACAGTGGTTCCTCGTCACGGACTGGAAACCCCTGACCATCTGG 240	
QY	213 PheTyrLysGluSerTyrLeuArgPheSerThrGlnArgPheSerLeuAspLysLeuAsp 232	
DB	241 TTCCTACAAGGAGAGTTACTTGGCGTTCCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGAC 300	
QY	233 SerAlaIleHisLeuCysAsnAsnAlaValGlnLysTyrLeuLysAsnAspValGlyArg 252	
DB	301 AGCGCCATCCACCTGTGCAACACCGCGTCCAGAGTACATCGACAGCGCTGCTCATCTGTGAC 360	
QY	253 ThrLysPheAspIleArgGlnTrpPheLeuValThrAspTrpAsnProLeuThrIleTrp 212	
DB	361 AGCCCCCTGCTGCCGCACACATGTGGACCAAGTACCTGAAGATCTCCAGGAGTACCTGCGAG 420	
QY	273 ArgGlnGlyArgGlyAlaValTrpGlySerValIleTyrProSerMetLysLysAlaIle 292	
DB	421 CGCCAGGCGCGTGGCGCGTGTGGGCGAGCGTTCATCTACCCGTCATGAAGAAGGCCATC 480	
QY	293 AlaHisAlaMetLysValAlaGlnAspHisValGluProArgLysAsnSerPheGluLeu 312	
DB	481 GCCCACCCATGAAGGTGGCCAGGACCAAGTGGAGCTCGCAAGAACAGCTTTGAGCTC 540	
QY	313 TyrGlyAlaAspPheValLeuGlyArgAspPheArgProTrpIleIleGluIleAsnSer 332	
DB	601 AGCCCCACCATGACCCCGTCCAGCCGTCACGGCCAGCTGTGTGCACAGGTGCAGGAG 660	
QY	353 AspThrIleLysValAlaValAspArgSerCysAspIleGlyAsnPheGluLeuLeuTrp 372	
DB	661 GACACCATCAAGGTGGCGGTGGACCGCGAGCTGTGACATCGGCAACTTCGAGCTCCTGTGG 720	
QY	373 ArgGln 374	
DB	721 AGGCAG 726	
RESULT 13		
US-10-635-977-3		
; Sequence 3, Application US/10635977		
; Publication No. US20040171131A1		
; GENERAL INFORMATION:		
; APPLICANT: Bristol-Myers Squibb Company		
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN		
; FILE OF INVENTION: TYROSINE-LIGASE-LIKE PROTEIN, BGS42		
; FILE REFERENCE: D0283A CIP		
; CURRENT APPLICATION NUMBER: US/10/635,977		
; CURRENT FILING DATE: 2003-08-07		
; PRIOR APPLICATION NUMBER: U.S. 60/394,725		
; PRIOR FILING DATE: 2002-07-09		
; PRIOR APPLICATION NUMBER: U.S.10/615,659		
; PRIOR FILING DATE: 2003-07-09		

RESULT 14

US-10-275-595A-39

; Sequence 39, Application US/10275595A

; Publication No. US20040078804A1

; GENERAL INFORMATION:

; APPLICANT: YUE, Henry

; APPLICANT: TANG, Y. Tom

; APPLICANT: AU-YOUNG, Janice

; APPLICANT: LU, Dying, Alina M.

; APPLICANT: BAUGHN, Maria R.

; APPLICANT: HILLMAN, Jennifer L.

; APPLICANT: AZIMZAI, Yalda

; APPLICANT: LAL, Preeti

; APPLICANT: YAO, Monique G.

; APPLICANT: BANDMAN, Olga

; APPLICANT: BURFORD, Neil

; APPLICANT: BATRA, Sajeev

; APPLICANT: KEARNEY, Liam

; APPLICANT: POLICKY, Jennifer L.

; TITLE OF INVENTION: CYTOSKELETON-ASSOCIATED PROTEINS

; FILE REFERENCE: PF-0772 USN

; CURRENT APPLICATION NUMBER: US/10/275,595A

; CURRENT FILING DATE: 2003-06-13

; PRIOR APPLICATION NUMBER: US 60/201,960

; PRIOR FILING DATE: 2000-05-05

; PRIOR APPLICATION NUMBER: US 60/202,729

; PRIOR FILING DATE: 2000-05-08

; PRIOR APPLICATION NUMBER: US 60/209,705

; PRIOR FILING DATE: 2000-06-05

; PRIOR APPLICATION NUMBER: US 60/210,149

; PRIOR FILING DATE: 2000-06-07

; PRIOR APPLICATION NUMBER: US 60/213,215

; PRIOR FILING DATE: 2000-06-21

; NUMBER OF SEQ ID NOS: 68

; SOFTWARE: PERL Program

; SEQ ID NO 39

; LENGTH: 2380

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc feature

; OTHER INFORMATION: Incyte ID No: 2156553CB1

US-10-275-595A-39

Alignment Scores:

Pred. No.:	1,15e-100	Length:	2380
Score:	1070.50	Matches:	250
Percent Similarity:	54.8%	Conservative:	81
Best Local Similarity:	41.4%	Mismatches:	198
Query Match:	37.5%	Indels:	76
DB:	7	Gaps:	12

US-10-635-977-2 (1-541) x US-10-275-595A-39 (1-2380)

QY 2 AlaSerSerIleLeuLysTrpValValSerHisGlnSerCysSerArgSerSerArgSer 21

DB 285 GCCGCAACGTTCTCAAGCTGGTGGTGAAGTCTGAG-----TGAAGTCA 329

QY 22 LysProArgAspGlnArgGluAlaGlySerSerAspLeuSerSerArgGlnAspAla 41

DB 330 TACCTATTTCAGGCGAGTAGAGAGAGCGCTTCAGAGACAAGCCCAAGAAACAGGAG 389

QY 42 GluAsnAlaGluAlaLysLeuArgGlyLeuProGlyGlnLeuValAspIleAlaCysLys 61

DB 390 AAAAACCCAGTGTTG-----GTGTCCCCAGAGTTGTGGATGAAGCTCTGTGT 437

QY 62 ValCysGlnAlaTyLeuGlyGlnLeuGluHisGluAspIleAspThrSerAlaAspAla 81

DB 438 GCGTGGAGGAGTACCTTAGCAACTTGGCCACATGGACATCGACAAGGACCTGGAGGCC 497

QY 82 ValGluAspLeuThrGluAlaGluTrpGluAspLeuThrGlnGlnTyLeuSerLeuVal 101

DB 498 CCGCTGTACCTACCCCGAGGGCTGGTCCCTCTTCCTCCAGCGCTACTACCAAGTGGTC 557

QY 102 HisGlyAspAlaPheIleSerAsnSerArgAsnTyLeuPheSerGlnCysGlnAlaLeuLeu 121

DB 558 CACGAAGGGCGAGAACTCAGGCACCTCGACACTCAGGTCCAGCGCTGTGAGGACATCCTG 617

QY 122 AsnArgIleThrSerValAsnProGlnThrAspIleAspGlyLeuArgAsnIleTrpIle 141

DB 618 CAGCAGCTGCAGGCGGTATCCCCAGATAGATGGAGGGGATCGCAACATCTGGATC 677

QY 142 IleLysProAlaAlaLysSerArgGlyArgAspIleValCysMetAspArgValGluGlu 161

DB 678 GTGAAGCCAGGAGCCCAAGTCCCGTGGAGGATCATGTGATGGACCATCTGGAGGAG 737

QY 162 IleLeuGluLeuAlaAlaAspHisProLeuSerArgAspAsnLysTrpValValGln 181

DB 738 ATGCTGAAGCTGGTGAACGGCAACCCCGTGGTGTGATGAAGACCGCAAGTGGGTGCAG 797

QY 182 LysTyIleGluThrProLeuLeuIleCysAspThrLysPheAspIleArgGlnTrpPhe 201

DB 798 AAGTATATTGAGCGGCGCTCTCTCATCTTGGCACCAAGTTTGACCTCAGACAGTGGTTC 857

QY 202 LeuValThrAspTrpAsnProLeuThrIleTrpPheTyLeuLysGluSerTyLeuArgPhe 221

DB 858 CTGGTAACCTGACTGGAACCCACTTACCGTGGTGTCTACCGGACAGCTATATCCGCTTT 917

QY 222 SerThrGlnArgPheSerLeuAspLysLeuAspSerAlaIleHisLeuCysAsnAsnAla 241

DB 918 TCCAGCGACGCTCTCTCCCTGAAGAACCTGGGACAACTCAGTGACACTGTGCAACAACCTCC 977

QY 242 ValGlnLysTyLeuLeuAsnAspValGlyArgSerProLeuLeuProAlaHisAsnMet 261

DB 978 ATCCAGAAAGCACCTGGAGAACTCATGCCATCGGCATCCACTGCTTCCGCCAGACAACATG 1037

QY 262 TrpThrSerThrArgPheGlnGluTrpLeuGlnArgGlnGlyArgGlyAlaValTrpGly 281

DB 1038 TGGTCTAGCAGAGGTTCCAGGCCCACTTCGAGGAGATGGGTGCCCAAACTCTTGGTCC 1097

QY 282 SerValIleTyProSerMetLysLysAlaIleAlaHisAlaMetLysValAlaGlnAsp 301

DB 1098 ACCATCATCGTGGCTGGCATGAAGATGCTGTGTATCCAGCACCTTCAGACCTCCCGAGAC 1157

QY 302 HisValGluProArgLysAsnSerPheGluLeuTyLeuGlyAlaAspPheValLeuGlyArg 321

DB 1158 ACCGTGCAGTGTCCGAAAGGCCAGCTTTGAGCTCTATGGCGCTGACTTGTGTTCGGGAG 1217

QY 322 AspPheArgProTrpLeuIleGluLeuAsnSerProThrMetHisProSerThrPro 341

DB 1218 GACTTCCAGCCCTGGCTGATTGAGATCAACCCAGCCCAAGATGGCACCTCTCCACAGCA 1277

QY 342 ValThrAlaGlnLeuCysAlaGlnValGlnGluAspThrIleLysValAlaVal----- 359

DB 1278 GTCACTGCCGGCTCTGTGTGGCTGCAAGCTGACACCTGCGGTGGTGTTCGACCGG 1337

QY 360 -----AspArgSerCysAspIleGlyAsnPheGluLeuLeuTrpArgGlnProValVal 377

DB 1338 ATGCTGGACCGCAACTGTGACACAGGAGCTTTGAGCTCATCTATAAGCAGCTCTGTG 1397

QY 378 GluProProProPheSerGlySerAspLeuCysValAlaGlyValSerValArgArg--- 396

DB 1398 GAGTGGCTCAATATGTGGGCATCCGGCTCTCTGTAGAGGGCTTCACCATCAAGAAGGCC 1457

QY 397 -----AlaArgArgGlnValLeuProValCysAsnLeu----- 407

DB 1458 ATGGCGATGTGTATCGCGGATGGGGGTTCGCCCCAGAGTCCCTCTCTGTACCCAGCGA 1517

QY 408 -----LysAlaSer-AlaSerLeuLeuAspAlaGlnProLeuLysAl 421

DB 1518 GGCTCTGGGGAGGCAAGGAGCTCGGGATCCCTACCCACAGGTCTTCTAGGAAAGGC 1577

QY 421 aArgGlyProSerAlaMetProAspProAlaGlnGlyProProSerProAlaLeuGln--- 440

DB 1578 ACTGGGGCCAGGAGGCTGGGGCAAGTGTCCACTGCTCCACTGCCACCACCTTCAGCC 1637

```
Qy 441 -ArgAspLeuGlyLeuLysGluGluLysGlyLeuProLeuAlaLeuLeuAlaProLeuAr 460
Db 1638 CCGGAAAGGGGAAGGCAAGCGAAAGGGCCACAGCCCTGGTCTGCCCAATCTC 1697
Qy 460 GGlyAlaAla-----GluSerGlyGlyAlaAlaGlnProThrArgThrLysAlaAl 477
Db 1698 TGGAGTGGGATGCCCCAGCAGCAGGATGGGCTGCATTTTC-ACCATGACCTTTTCTAG 1756
Qy 477 aGlyLysValGlu-----LeuProLacysProCysArgHisValAspSerGlnAl 494
Db 1757 TGGGGACAGCAACCCACCACCTTGAACAGATTGCCACTGAGTCCGAGAACCCCCAGGC 1816
Qy 494 aProAnThrGlyValProValAlaGln-----ProAlaLys 506
Db 1817 CTTGGGTAAACATTTCCCCCAAAACACCCGAGTGTTCAAAGCGGATTTATTCTCTCT 1876
Qy 506 sSerTrpAspProAnGlnLeuAsnAlaHisPro----- 517
Db 1877 CCAAGCCCTCCCAACCACTGGATCAGCCACCCCAAGAGCCACCAAGTAGCAAGTA 1936
Qy 517 ----- 517
Db 1937 AAAGCCACTACTCACAAGTATTGTTTAAATAATACACAGCCAAATTAGCTGGGCGCGTG 1996
Qy 518 -LeuGluProValLeuArgGlyLeuLysThrAlaGluGlyAlaLeuArgProProGln 537
Db 1997 GTGTGAGCCTGTGTCCAGGCTACTCGCGAGGCTAATGAGGAT---CGTTGAGCCCGG 2053
Qy 537 YGlyLysGly 540
Db 2054 AGGTCAAGGC 2063
```

RESULT 15

```
US-10-102-524-1765
; Sequence 1765, Application US/10102524
; Publication No. US20030109434A1
; GENERAL INFORMATION:
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Gaiger, Alexander
; APPLICANT: Gordon, Brian
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TREATMENT OF KIDNEY DISEASE
; FILE REFERENCE: 210121.572
; CURRENT APPLICATION NUMBER: US/10/102,524
; CURRENT FILING DATE: 2002-03-19
; NUMBER OF SEQ ID NOS: 1863
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 1765
; LENGTH: 2538
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-102-524-1765
```

Alignment Scores:

```
Pred. No.: 2,32e-95 Length: 2538
Score: 1020.00 Matches: 220
Percent Similarity: 55.5% Conservatives: 69
Best Local Similarity: 42.2% Mismatches: 151
Query Match: 35.7% Indels: 81
DB: 5 Gaps: 12
```

US-10-635-977-2 (1-541) x US-10-102-524-1765 (1-2538)

```
Qy 54 GlnLeuValAspLeuAlaCysLysValCysGlnAlaTrpLeuGlyGlnLeuGluHisGlu 73
Db 835 GAGTTTGTGATGAAGCTCTGTGTGCTGCGGAGGAGTACCTTAGCAACTTTGGCCCATG 894
Qy 74 AspLeuAspThrSerAlaAspAlaValGluAspLeuThrGluAlaGluTrpGluAspLeu 93
Db 895 GACATCGACCAAGNACTGGAGGCCCTGCTACCTACCCCGAGGCGCTGGTCCCTCTTC 954
```

```
Qy 94 ThrGlnGlnTyrTyrSerLeuValHisGlyAspAlaPheIleSerAsnSerArgAsnTyr 113
Db 955 CTCACAGCGCTACTACCAAGTGTCTCCAGNAGGGGAGAACTCAGGACCTCCGACACTCAG 1014
Qy 114 PheSerGlnCysGlnAlaLeuLeuAsnArgIleThrSerValAsnProGlnThrAspLeu 133
Db 1015 GTCCAGCGCTGTGAGGACATCTCTGAGCAGCTGCAGGCGGTGTACCCCGAGATAGACATG 1074
Qy 134 AspGlyLeuArgAsnIleTrpIleLeuLysProAlaAlaLysSerArgGlyArgAspLeu 153
Db 1075 GAAGGGGATCGCAACATCTGGATCTGTAAGCAGGAGGCCAAGTCCCGCGGAGGAGCATC 1134
Qy 154 ValCysMetAspArgValGluGluLeuAlaAlaAspHisProLeuSer 173
Db 1135 ATGTGATGAGACCACTGGAGGAGATCTGAAGCTGGTGAACGGCAACCCCTGGTGTATG 1194
Qy 174 ArgAspAsnLysTrpValValGlnLysTyrIleGluThrProLeuLeuLeuLeuCysAspThr 193
Db 1195 AAGGACGGCAAGTGGGTGGTGCAGAAATATATTGAGCGGCCCTCTCTCATCTTTTGGCACC 1254
Qy 194 LysPheAspIleArgGlnTrpPheLeuValThrAspTrpAsnProLeuThrIleTrpPhe 213
Db 1255 AAGTTTACCTCAGACAGTGTCTCTGTAACCTGAGTGAACCCACTTACCGTGTGGTTC 1314
Qy 214 TyrLysGluSerTyrLeuArgPheSerThrGlnArgPheSerLeuAspLysLeuAspSer 233
Db 1315 TACCGGAGCAGCTATATCCGCTTTTCCACGAGCCCTTCTCTCTGAGAACTTGGACACAC 1374
Qy 234 AlaIleHisLeuCysAsnAsnAlaValGlnLysTyrLeuLysAsnAspValGlyArgSer 253
Db 1375 TCAGTCACCTGTGCAACAACTCCATCCAGAGAACCTCGAGAACTCATGCCATCGGCAT 1434
Qy 254 ProLeuLeuProAlaHisAsnMetTrpThrSerThrArgPheGlnGluTrpLeuGlnArg 273
Db 1435 CCACCTGCTTCCGCCAGACAAACATGTGTCTAGCCAGAGTTCCAGGCCCTCCAGGAG 1494
Qy 274 GlnGlyArgGlyAlaValTrpGlySerValIleTyrProSerMetLysLysAlaIleAla 293
Db 1495 ATGGGTGCCCAATGCTTGGTCCACCATCATCGTCCCTGGCATGAGAGATGCTGTGATC 1554
Qy 294 HisAlaMetLysValAlaGlnAspHisValGluProArgLysAsnSerPheGluLeuTyr 313
Db 1555 CAGGCATTTCCAGACCTCCCGAGGAGACCGTGCAGTGTGCGAAGGCCAGCTTTGAGCTCAT 1614
Qy 314 GlyAlaAspPheValLeuGlyArgAspPheArgProTrpIleGluLeuLeuAsnSerSer 333
Db 1615 GCGCTGACTTCGTGTTCGGGAGGACTTCCAGCCCTGCTGATTCAGATCAACAGCCAGC 1674
Qy 334 ProThrMetHisProSerThrProValThrAlaGlnLeuCysAlaGlnValGlnGluAsp 353
Db 1675 CCCACGATGGCACCTCCACAGCAGTCACTGCGCGCTCTGTGCTGGCGTGCAGCTGAC 1734
Qy 354 ThrIleLysValAlaVal-----AspArgSerCysAspIleGlyAsnPheGlu 369
Db 1735 ACCCTGCGGTGTCTATTGACCGGATGCTGGACCGCAACTGTGACAGAGGCTTTGAG 1794
Qy 370 LeuLeuTrpArgGlnProVal---ValGluPro-ProProPheSerGlySerAspLeuCy 388
Db 1795 CTCATCTATAAGCAGCCCGCTCACCACTTCCCGAGCTCCACACCAAGGCC----- 1843
Qy 388 sValAlaGlyValSerValArgAlaArgGlnValLeuProValCysAsnLeuLys 408
Db 1843 ----- 1843
Qy 408 sAlaSerAlaSerLeuLeuAspAlaGlnProLeuLysAlaArgGlyProSerAlaMetPr 428
Db 1844 -----CCAGCTGCTCTTCCCCCATGTACTCCGACACAGGCGCCAGGCTCTCAGACGACAGC 1899
Qy 428 oAspProAlaGlnGlyProProSerProAlaLeuGlnArgAspLeuGlyLysGluGln 448
Db 1900 ACAGCAAGCTGTGTGGGCACTTAAGGCCCTGTGCGACCAACAGCAAA---GGCCTTGGAGCTC 1956
Qy 448 Lys-----GlyLeuProLeuAlaLeuAlaPro---LeuArgGlyAla----- 462
```

```
Db      1957 TACCCACGGCTAAGGCTTCATTTCCTCCACGAACCTTGATTTCAAGGTGGCACCCA 2016
Qy      |||||
Qy      463 -----AlaGluSerGlyAlaAlaGlnProThrArgThr----- 474
Db      |||::: |||||
Db      2017 GCATCCTGAAGCCAAGAAAGGTGGCTCGACCTGTGACTCACACCCAGTGGACAGTGCT 2076
Qy      |||
Qy      475 -----LysAlaAlaGlyLysValGluLeuProAlaCysProCysAr 488
Db      |||::: |||||
Db      2077 GAGCACGGGTCAAGGCTGAGGGGCACAGGCAGAGGGCAGCTCCAGGCTGG----- 2128
Qy      488 gHisValAspSerGlnAlaProAsnThrGlyValProValAlaGlnProAlaLysSerTr 508
Db      |||::: |||||
Db      2129 -----CTGGCACCCCAAGGGAAGAGCTG 2151
Qy      508 pAspProAsnGlnLeuAsn-----AlaHisProLeuGluProValLe 522
Db      |||::: |||||
Db      2152 GTCTCCCTCAGAAGCCCTTCCTCCACAGACTTCTGATCATCTCCCTCTCTCCCTCCT 2211
Qy      522 uArgGlyLeuLysThrAlaGluGlyAlaLeuArgPro-----ProProGlyGl 538
Db      |||::: |||||
Db      2212 T-----TCACACCGAGGCTCCTGCTCTCCTGTGCTTCCGAGGCCCCCAGCTGG 2259
Qy      538 Y 538
Db      2260 A 2260
```

Search completed: April 4, 2006, 11:54:48
Job time : 1232 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 4, 2006, 09:38:21 ; Search time 288 Seconds
(without alignments)
3339.100 Million cell updates/sec

Title: US-10-635-977-2

Perfect score: 2854

Sequence: 1 MASSILKVVVSHQSCRSR.....LRGLKTAEGALRPPPGKGKS 541

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-O=/abs/ABSSWEB spool/US10635977/runat_04042006.103535.10086/app_query.fasta_1
-DB=Issued Patents NA -QFMT=fastap -SUFFIX=p2n.rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=Bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-HOST=abs02h -USER=US10635977 @CGN 1 1.193 @runat 04042006.103535.10086
-NCPU=6 -ICPU=3 -NO MMAP -NEG SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WREN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents NA:
1: /cgn2_6/ptodata/1/ina/1 COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5 COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/1/ina/H COMB.seq.*
6: /cgn2_6/ptodata/1/ina/PCTUS COMB.seq.*
7: /cgn2_6/ptodata/1/ina/PP COMB.seq.*
8: /cgn2_6/ptodata/1/ina/RE COMB.seq.*
9: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1020	35.7	3001	3	US-09-620-312D-494
2	305	10.7	5314	3	US-09-620-312D-138
3	302	10.6	2624	3	US-10-104-047-1688
4	271.5	9.5	2074	3	US-10-104-047-1401
5	228.5	8.0	1980	3	US-09-270-767-15138
6	210	7.4	2111	3	US-10-104-047-1651
7	205.5	7.2	2716	3	US-09-799-451-295
8	166	5.8	3394	3	US-09-814-915A-56
9	154.5	5.4	783	3	US-09-270-767-31557

10	154.5	5.4	1525	3	US-09-270-767-15222	Sequence 15222, A
11	153	5.4	30135	3	US-09-902-540-1249	Sequence 1249, Ap
12	151	5.3	780	3	US-09-248-796A-500	Sequence 500, App
13	132	4.6	3027	3	US-09-252-991A-12848	Sequence 12848, A
14	129.5	4.5	1212	3	US-09-252-991A-12504	Sequence 12504, A
C 15	129.5	4.5	1251	3	US-09-252-991A-12970	Sequence 12970, A
C 16	129.5	4.5	7419	3	US-09-252-991A-481	Sequence 481, App
C 17	129.5	4.5	7449	3	US-09-252-991A-396	Sequence 396, App
C 18	129	4.5	1188	2	US-08-926-253-1	Sequence 1, Appli
C 19	129	4.5	1188	2	US-08-926-253-3	Sequence 1, Appli
C 20	129	4.5	1188	3	US-09-120-074-1	Sequence 1, Appli
C 21	129	4.5	1188	3	US-09-120-074-3	Sequence 3, Appli
C 22	129	4.5	1297	3	US-09-270-767-15012	Sequence 15012, A
C 23	129	4.5	72704	3	US-09-902-540-1273	Sequence 1273, Ap
C 24	127	4.4	2534	3	US-09-096-776B-1	Sequence 1, Appli
C 25	127	4.4	2534	3	US-09-923-922-1	Sequence 1, Appli
C 26	127	4.4	3765	3	US-09-949-016-5816	Sequence 5816, Ap
C 27	127	4.4	3801	3	US-09-949-016-907	Sequence 907, App
C 28	127	4.4	23951	3	US-09-902-540-1245	Sequence 1245, Ap
C 29	127	4.4	27903	3	US-09-902-540-1235	Sequence 1235, Ap
C 30	126.5	4.4	12194	3	US-09-902-540-1091	Sequence 1091, Ap
C 31	126	4.4	2715	3	US-09-252-991A-12548	Sequence 12548, A
C 32	126	4.4	5173	3	US-09-949-016-11822	Sequence 11822, A
C 33	126	4.4	5173	3	US-09-949-016-13985	Sequence 13985, A
C 34	125.5	4.4	3342	3	US-09-252-991A-12644	Sequence 12644, A
C 35	125.5	4.4	5467	3	US-09-902-540-703	Sequence 703, App
C 36	125.5	4.4	6244	2	US-08-076-726-15	Sequence 15, Appli
C 37	125.5	4.4	6244	2	US-08-260-452-8	Sequence 8, Appli
C 38	125.5	4.4	6244	2	US-08-481-970-8	Sequence 8, Appli
C 39	125.5	4.4	6244	2	US-08-897-719-8	Sequence 8, Appli
C 40	125.5	4.4	6244	3	US-09-163-269-8	Sequence 8, Appli
C 41	125.5	4.4	6244	3	US-09-281-674-8	Sequence 8, Appli
C 42	125.5	4.4	6244	3	US-09-777-317B-8	Sequence 8, Appli
C 43	125.5	4.4	7305	3	US-09-902-540-961	Sequence 961, App
C 44	125.5	4.4	61158	3	US-09-949-016-15041	Sequence 15041, A
C 45	125.5	4.4	75431	3	US-09-949-016-15122	Sequence 15122, A

ALIGNMENTS

RESULT 1

US-09-620-312D-494
; Sequence 494, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yunqing
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John Tillinghast
; APPLICANT: Dmanac, Radoje T.
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 784CIP2B
; CURRENT APPLICATION NUMBER: US/09/620,312D
; CURRENT FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1105
; SOFTWARE: pt_FL_genes Version 1.0

```
; SEQ ID NO 494
; LENGTH: 3001
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1361)..(2419)
US-09-620-312D-494

Alignment Scores:
Pred. No.: 7,34e-84 Length: 3001
Score: 1020.00 Matches: 220
Percent Similarity: 55.5% Conservative: 69
Best local Similarity: 42.2% Mismatches: 151
Query Match: 35.7% Indels: 81
DB: 3 Gaps: 12

US-10-635-977-2 (1-541) x US-09-620-312D-494 (1-3001)

Qy 54 GlnLeuValAspIleAlaCysLysValCysGlnAlaTyrLeuGlyGlnLeuGluHisGlu 73
Db ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
1304 GAGTTTGTGATGAAGCTCTGTGTGCTGCGAGAGTACCTTAGCAACTTTGGCCCATG 1363
Qy 74 AspIleAspThrSerAlaAspAlaValGluAspLeuThrGluAlaGluTTPGluAspLeu 93
Db ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
1364 GACATCCAGCAAGGACCTGGAGGCCCTGCTACTACCCCGAGGGCTGGTCCCTCTTC 1423
Qy 94 ThrGlnGlnTyrTyrSerLeuValHisGlyAspAlaPheIleSerAsnSerArgAsnTyr 113
Db ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
1424 CTCAGGGCTACTACCAAGTGGTCCACGAAGGGCGAGAATCTCAGGCACCTCGACACTCAG 1483
Qy 114 PheSerGlnCysGlnAlaLeuLeuAsnArgIleThrSerValAsnProGlnThrAspIle 133
Db ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
1484 GTCCAGCGCTGTGAGGACATCTGCAGCAGCTGCAGCGCTGTGTACCCAGATAGACATG 1543
Qy 134 AspGlyLeuArgAsnIleTTPleIleLysProAlaAlaLysSerArgGlyArgAspIle 153
Db ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
1544 GAAGGGATCGCAACATCTGGATCGTGAAGCCAGGAGCAAGTCCCGGACGAGGATC 1603
Qy 154 ValCysMetAspArgValGluGluIleLeuGluLeuAlaAlaAspHisProLeuSer 173
Db ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
1604 ATGTGCATGACCCACCTGGAGAGATGCTGAAGTGTGAACGGCACCCCGTGTGATG 1663
Qy 174 ArgAspAsnLysTTPValValGlnLysTyrIleGluThrProLeuLeuIleCysAspThr 193
Db ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
1664 AAGGACGGCAAGTGGTGGTGCAGAGATATATTGAGCGGCCCTCTCATCTTTGGCACC 1723
Qy 194 LysPheAspIleArgGlnTTPheLeuValThrAspTTPAsnProLeuThrIleTTPhe 213
Db ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
1724 AAGTTTGACCTCAGACAGTGTCTCTGTGAATGACTGGAACCCACTTACCGTGTGTTTC 1783
Qy 214 TyrLysGluSerTyrLeuArgPheSerThrGlnArgPheSerLeuAspLysLeuAspSer 233
Db ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
1784 TACCGCAGACGATATATCCGCTTTTCCACGAGCCCTCTCCCTGGAAGAACCTGGACAAC 1843
Qy 234 AlaIleHisLeuCysAsnAsnAlaValGlnLysTyrLeuLysAsnAspValGlyArgSer 253
Db ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
1844 TCAGTGACCTGTGCAACAACTCCATCCAGAAGCACCTGGAGAACTCATGTCATCGGCAT 1903
Qy 254 ProLeuLeuProAlaHisAsnMetTTPThrSerThrArgPheGlnGluTyrLeuGlnArg 273
Db ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
1904 CCATGTCTCGCCAGACAAACATGTGTGTAGCCAGAGTTTCCAGGCCACCTCGAGAG 1963
Qy 274 GlnGlyArgGlyAlaValTTPGlySerValIleTyrProSerMetLysLysAlaIleAla 293
Db ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
1964 ATGGGTGCCCCAAATGTGTGGTCCACCATCATCTGCTGCGCATGAAGGATGCTGTGATC 2023
Qy 294 HisAlaMetLysValAlaGlnAspHisValGluProArgLysAsnSerPheGluLeuTyr 313
Db ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
2024 CACGCACTTCAGACTCCAGGACACCGTGCAGTGTGGAAGGCCAGCTTTGACCTCTAT 2083
Qy 314 GlyAlaAspPheValLeuGlyArgAspPheArgProTTPLeuIleGluIleAsnSerSer 333
Db ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
```

RESULT 2

```
US-09-620-312D-138
; Sequence 138, Application US/09620312D
; Patent No. 659662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing-A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yunqing
; APPLICANT: Wang, Dunrui
```


; APPLICANT: Wang, Zhiwei
 ; APPLICANT: John Tillinghast
 ; APPLICANT: Dmanac, Radoje T.
 ; TITLE OF INVENTION: No. 656962el Nucleic Acids and
 ; FILE REFERENCE: Polypeptides
 ; CURRENT APPLICATION NUMBER: 784CIP2B
 ; CURRENT FILING DATE: 2000-07-19
 ; PRIOR APPLICATION NUMBER: US/09/620,312D
 ; PRIOR FILING DATE: 2000-04-25
 ; PRIOR APPLICATION NUMBER: 09/488,725
 ; PRIOR FILING DATE: 2000-01-21
 ; NUMBER OF SEQ ID NOS: 1105
 ; SOFTWARE: pt_FL_genes Version 1.0
 ; SEQ ID NO 138
 ; LENGTH: 5314
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (531)..(4376)
 ; US-09-620-312D-138

Alignment Scores:
 Pred. No.: 1.89e-17 Length: 5314
 Score: 305.00 Matches: 89
 Percent Similarity: 49.3% Conservative: 50
 Best Local Similarity: 31.6% Mismatches: 93
 Query Match: 10.7% Indels: 51
 DB: 3 Gaps: 11

US-10-635-977-2 (1-541) x US-09-620-312D-138 (1-5314)

QY 137 ArgAsnileTrpIlelelyysProAlaAlaLysSerArgGlyArgAspileValCysMet 156
 DB 1050 CGGGACCTTGGATAGTAAACCAAGTCATCTTCAAGGGGGGGGGCTACCTGATC 1109
 QY 157 AspArgValGluGluLeuGluLeuAlaAlaAspHisProLeuSerArgAspAsn 176
 DB 1110 AACAAATCCAAACCATC-----TCCCTGGAAGAG 1139
 QY 177 LysTrpValValGlnLysTyrIleGluThrProLeuLeuLeuCysAspThrLysPheAsp 196
 DB 1140 AACATTTGGTCCCGTTACATTAACACCCCTGCTCATAGATGATTTCAAGTTTGAC 1199
 QY 197 IleArgGlnTrpPheLeuValThrAspTrpAsnProLeuThrIleTrpPheTyrLysGlu 216
 DB 1200 GTGGCCCTCTAGTGTCTGTGACTTCCATGATCTCTTGTCTATCTATCTATGAGAA 1259
 QY 217 SerTyrLeuArgPheSerThrGlnArgPhe-----SerLeuAspLysLeuAspSer 233
 DB 1260 GGATTTGGCTAGGTTTGGCAACTGTGGATATGATCAAGAGCCCAAGACATTCCGGAACCA 1319
 QY 234 AlalleHisLeuCysAsnAsnAlaValGlnLys-----TyrLeuLysAsnAsp 249
 DB 1320 TTCATGCTCTGACAACTACAGTGTCAACAGAAAGAGTGGAGATTACGTGATTTGTGAC 1379
 QY 250 ValGlyArgSerProLeuProAlaHis---AsnMetTrpThrSerThrArgPheGln 268
 DB 1380 -----GATCCAGAGTGGAGGATTATGGAACAATGGAGCATGCTATGCTT 1430
 QY 269 GluTyrLeuGlnArgGlnGlyArgGlyAla----- 278
 DB 1431 AGGTACCTGAAACAGAGAGGAGAGATACACCCGATTTGATGGCCCATGTAGAACACCTG 1490
 QY 279 ValTrpGlySerValIleTyrProSerMetLysLysAlaIleAlaHisAlaMetLysVal 298
 DB 1491 ATCACTAAGACTATAATC-----TCTGCTGAACCTAGCTATTTGCTGCTGCTGTAACACC 1544
 QY 299 AlaGlnAspHisValGluProArgLysAsnSerPheGluLeuTyrGlyAlaAspPheVal 318
 DB 1545 TTTGTTCTCAT-----CGCAGCAGTTGTTTTGAACTCTATGGCTTTGACGTGCTC 1595

QY 319 LeuGlyArgAspPheArgProTrpLeuIleGluLeuAsnSerSerProThrMetHisPro 338
 DB 1596 ATAGATTCTACTCTGAAGCCATGGTGTGTGAAGTGAATCTCTCTCTTTTGGCCTGT 1655
 QY 339 SerThrProValThrAlaGlnLeuCysAlaGlnValGlnGluAspThrIle----- 355
 DB 1656 GATGCGCTCTGGACCTAAAGATTAAAGCCATGATGATTTCAGATAT-GTTCACTCTTGT 1714
 QY 356 LysValAlaValAspArgSerCysAspileGly-----Asn-PH 368
 DB 1715 AGGATTTGTGTCGAAGATCTCTGCCAGCGGGCATCACTCGGCCAATTTATCCACCTT 1774
 QY 368 eGluLeuLeuTrpArgGlnProValValGluProPro-----ProPheSerGlySe 385
 DB 1775 CGAGTCTTCCAGCGAAACCTTTCAGAAACCTCAGCGTTCGCGTCCACTCTCTGCCAG 1834
 QY 385 rAsp 386
 DB 1835 TGAT 1838
 RESULT 3
 US-10-104-047-1688
 ; Sequence 1688, Application US/10104047
 ; Patent No. 6943241
 ; GENERAL INFORMATION:
 ; APPLICANT: HELIX RESEARCH INSTITUTE
 ; TITLE OF INVENTION: No. 6943241el full length cDNA
 ; FILE REFERENCE: H1-A0105
 ; CURRENT APPLICATION NUMBER: US/10/104,047
 ; CURRENT FILING DATE: 2002-03-25
 ; PRIOR APPLICATION NUMBER:
 ; PRIOR FILING DATE:
 ; NUMBER OF SEQ ID NOS: 4096
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 1688
 ; LENGTH: 2624
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; US-10-104-047-1688
 Alignment Scores:
 Pred. No.: 1.19e-17 Length: 2624
 Score: 302.00 Matches: 132
 Percent Similarity: 37.7% Conservative: 76
 Best Local Similarity: 24.0% Mismatches: 200
 Query Match: 10.6% Indels: 144
 DB: 3 Gaps: 23
 US-10-635-977-2 (1-541) x US-10-104-047-1688 (1-2624)
 QY 82 ValGluAspLeuThrGluAlaGluTrpGlu-----GlnGlnTyr 97
 DB 447 GTGAAGTGAAGGAGGAGGGAGTGGATTTCTACTGTGTGACGTGAGTGGCTCGG 506
 QY 92 -----AspLeuThr----- 91
 DB 507 GAGAACTTCGACCACCATCATCGATGCAATGTCGGATCAGTCTTCGGAACCA 566
 QY 98 TyrSerLeuValHisGlyAspAlaPheIleSerAsnSerArgAsnTyrPheSerGlnCys 117
 DB 567 TATGAGCTGACCCCGAAGAACTATCATGTGTGAAGAACCTGAAACCGGTTCGGAAGCAGCTG 626
 QY 118 GlnAlaLeuLeuAsnArgIleThrSerValAsn-----ProGlnThr----- 131
 DB 627 GAGCGTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 686
 QY 132 -----AspIleAspGlyLeuArgAsn-----IleTrp 140
 DB 687 ATGCTTCCGAGTACCACCTGTTTGTAGAGGAGTTTCGAAAAAACCAGGAATCACCTGG 746
 QY 141 IleIleLysProAlaAlaLysSerArgGlyArgAspIleValCysMetAspArgValGlu 160
 DB 747 ATCATGAAGCCTGTAGCCCGGTCTCAAGGGAAGGAGCATCTTCTCTCCGTAGGCTGAAG 806

QY	161	GluileuLeuGluLeuAlaAlaAspHisProLeuSerArgAspAsnLys-----	177	QY	471	ProThrArgThrLysAlaAlaGlyLysValGluLeuProAlaCysProCysArgHisVal	490
Db	807	GACATCGTGAGCTGG---AGGAAGACACAGAAGCTCTGACGACCACAGAAAGATGATATT	863	Db	1787	CTTCTGCTTCCAGAGCATGAGGACTTCTCCACATCATCCCTGTACTTGAGGCACACACCA	1846
QY	178	-----TrpValValGlnLysTyrIleGluThrProLeuLeuLeuCysAspThr	193	QY	491	AspSerGlnAlaProAsnThrGlyValProValAlaGlnProAlaLysSerTrpAspPro	510
Db	864	CCCGTGAGAACTATGTGGCTCAGCGTTACATGAAATCTTTACCTGATGAGGCCGCG	923	Db	1847	CCAGACAG-----GTCCAGGCATGGACTCCA	1873
QY	194	LysPheAspIleArgGlnTrpPheLeuValThrAspTrpAsnProLeuThrIleTrpPhe	213	QY	511	AsnGlnLeuAsnAlaHisProLeuGluProValLeuArgGlyLeuLysThrAlaGluGly	530
Db	924	AAATTTGACCTGCTGCTATGCTGGTGAATGCTG-----GTG	962	Db	1874	GGTTCCTA-----TTTCCAGAGGAATGCGATTAGCTTTT	1912
QY	214	TyrLysGluSerTyrLeuArgPheSerThrGlnArgPheSerLeuAspLysLeuAspSer	233	QY	531	AlaLeuArgProPro---ProGlyGlyLysGly	540
Db	963	TTTGCTGAATGCTGCTGTGGTCTGGGCACAGGAGA-----CAG	1001	Db	1913	GGCCACGACCCACCCTGCCAGGAGGAGGGG	1945
QY	234	AlaIleHisLeuCysAsnAlaValGlnLysTyrLeuLysAsnAspValGlyArgSer	253	RESULT 4			
Db	1002	GATGTTCACTCACCAACGTGGCTGTGCMAAAA-----MetTrpThrSerThrArgPheGln	268	US-10-104-047-1401			
QY	254	ProLeuLeuProAlaHisAsn-----MetTrpThrSerThrArgPheGln	287	; Sequence 1401, Application US/10104047			
Db	1035	---ACATCTCCGACTACCAACCAAGAGGGCTGCAAGTGGAGCGCTGCGCTCCGG	1091	; Patent No. 6943241			
QY	269	GluTyrLeu---GlnArgGlnGlyArgGlyAlaValTrpGlySerValIleTyrProSer	287	; GENERAL INFORMATION:			
Db	1092	CAGTACTGGCGTCCAAACACGGCCGCGAGCAGTG-----GAGACACTCTTTCAGGGAC	1145	; APPLICANT: HELIX RESEARCH INSTITUTE			
QY	288	MetLysLysAlaIleAlaHisAlaMetLysValAlaGlnAspHisValGluProArgLys	307	; TITLE OF INVENTION: No. 6943241el full length cDNA			
Db	1146	ATCGACAACTCTTTGTCAAAAGCTCGAGAGTGTGCAGAAGGTGATCATCAGTGACAAG	1205	; FILE REFERENCE: H1-A0105			
QY	308	AsnSerPheGluLeuTyrGlyAlaAspPheValLeuGlyArgPheArgProTrpLeu	327	; CURRENT APPLICATION NUMBER: US/10/104,047			
Db	1206	CACTGCTTCGAGCTGTCAGGTATGATCATCTCATGACGACGACCTCAAGCCGTGGCTC	1265	; CURRENT FILING DATE: 2002-03-25			
QY	328	IleGluIleAsnSerProThrMetHisProSerThrProValThrAlaGlnLeuCys	347	; PRIOR APPLICATION NUMBER:			
Db	1266	CTGAGGTCAATGGTCTCCCATCTACAGCCAGCAGCCAGGAAGACTATGAGCTCAAG	1325	; PRIOR FILING DATE:			
QY	348	AlaGlnValGlnGluAspThrIleLysValAla-----ValAspArg	361	; NUMBER OF SEQ ID NOS: 4096			
Db	1326	ACCTGCCTCTCGGAAGACACCTGCATGTTGTGACATGGAAGCGAGGCTCACGGGAAG	1385	; SOFTWARE: PatentIn Ver. 2.1			
QY	362	SerCysAspIleGlyAsnPheGluLeuLeuTrpArgGlnProValValGlu-----	378	; SEQ ID NO 1401			
Db	1386	GAGAAGCGAGTCGGGGGCTTTGACCTCATGTGGAAATGATGGCCCTTTAGCAGAGAGAG	1445	; LENGTH: 2074			
QY	379	-----ProProProPheSerGlySerAspLeuCysValAlaGlyValSerValArgArg	396	; TYPE: DNA			
Db	1446	GGGGCTCTGA-CCTGTGCGGAATGGGAAACTTTGTGACCAACACACATCTCGGTGGT	1504	; ORGANISM: Homo sapiens			
QY	397	AlaArgArg-----	399	US-10-104-047-1401			
Db	1505	CAACGATCGGAAGAAACAACCTGAGCAGCTCTTCTGCTCCCTTCAAGTTCAGAAAGAAC	1564	Alignment Scores:			
QY	400	---GlnValLeuProValCysAsnLeuLysAlaSerAlaSerLeuLeuAspAlaGlnPro	418	Pred. No.: 5.42e-15			
Db	1565	TTCCAGTTGATCCCGCTGCCAGGAGGAATCAGCCTTAGC-----AGGCCT	1612	Score: 271.50			
QY	419	Leu-----LysAlaArgGlyProSerAlaMetProAspPro	430	Percent Similarity: 36.6%			
Db	1613	TGTGTTTGAGATTGGAGCTGAGCCAGAAACCGGAAACCTCTGGGCTCATGGCAGGACA	1672	Best Local Similarity: 23.6%			
QY	431	AlaGlnGlyProProSerProAlaLeuGlnArgAspLeuGlyLeuLysGluLysGly	450	Query Match: 9.5%			
Db	1673	GCTTCGGAGTTGAGTGTCAAGCCGAAAGGGAAGACTTT---TGGAAAGAAATACTGCG	1729	DB: 15			
QY	451	LeuProLeuAlaLeuLeuAlaProLeuArgGlyAlaAlaGluSerGlyGlyAlaAlaGln	470	US-10-635-977-2 (1-541) x US-10-104-047-1401 (1-2074)			
Db	1730	AAATCCAAAGAGTGAACCGACCGACTTT---TAGGAAAGCCAGAACCGAGCTACTCTAGA	1786	QY 2 AlaSerSerIleLeuLysTrpValValSerHisGlnSerCysSerArgSerSerArgSer			
				Db 14 GCCTCCGCGCATTTTCAGCTGGCGCTGCAGC-----TGCTGCACAGAGACCCACAGAGGC			
				QY 22 LysProArgAspGlnArgGluGluAlaGlySerSerAspLeuSerSerArgGln-----			
				Db 68 CACCTCGGAACCAACGCGCCCAATGAGAGGGCGGAGCTGTGTTCTCCACAAAGCCAG			
				QY 40 -----AspAla			
				Db 128 GCGCTGGATCTTTGAGAACCCACCCAGCCTTTACCTTAACATTCATCCAGGACA			
				QY 42 GluAsnAlaGluAlaLysLeuArgGlyLeuProGlyGlnLeuValAspIleAlaCysLys			
				Db 188 AACCACTGAGCAGCCGCTGCAGGCTGGGAGCAAGGCTACAGGAAGCAGGTGTTTCC			
				QY 62 Val-----			
				Db 248 ATCCCTCCAGCGGAGCGCCCAACACCAACACTGGAGAGAGAGAAAAAACCTCATTTG			
				QY 63 -----CysGlnAlaTyrLeuGlyGlnLeuGluHisGluAspIleAsp			
				Db 308 ATGGCGGAAGATGAACCTTCAGGGGCCCTCTTGAAGCCGCTGGTTTTCGCGTTGACGAG			
				QY 77 ThrSerAlaAspAlaValGluAsp-----LeuThrGluAlaGluTrpGluAspLeuThrGln			
				Db 368 ACCACCCCGGCTGTGGTGCMAAGTGTCTCTCTGGAGAGGGGGTGAATAAGTTTGATAAG			

```
QY 96 GlnTyr-----TyrSerLeuValHisGlyAspAlaPheIleSerAsnSer 110
|||
|||
Db 428 CAGGAGCAACCGGAGGACTGGAACTGTACTG-GAGGACATCTCTTCGGAATGAC 486
|||
|||
QY 111 ArgAsnTyrPheSerGlnCys-----
|||
|||
Db 487 CGRACA-----CAACAGTGTAAACCGTGGCAGCAGCTAAACCACCACCTGGGAACAC 540
|||
|||
QY 118 GlnAlaLeuAsnArgIleThrSerValAsnPro-----
|||
|||
Db 541 CAAGCTTACCAGGAAGACTGTTTGGCCAAACACCTGAAGCACATGAGGAGGATGTATGG 600
|||
|||
QY 129 -----
|||
|||
Db 601 CACTTCCCTGTACCACTTCATCCCTGACGTTGCTGATGCCCAATGACTATATACCAAGTT 660
|||
|||
QY 130 -----GlnThrAsp-IleAspGlyLeuArgAsnIleTrpIleI 142
|||
|||
Db 661 CGTGGCTGAATCTTTCAGGAGGAGGAGATGCTGGGCACCAAGCATAGCTATTTGATTGG 720
|||
|||
QY 142 eLysProAlaAlaLysSerArgGlyAspIleValCysMetAspArgValGluIuI 162
|||
|||
Db 721 CAAGCCTGCTGAGTTATCTCGTGGGAGGGGATACTAATTTTCAGTGACTTTAAAGACTT 780
|||
|||
QY 162 eLeuGluLeuAlaAlaAspHisProLeuSerArgAspAsnLysTrpValGlnIly 182
|||
|||
Db 781 CATC-----TTTGATGATATGTACATAGTCAGAA 810
|||
|||
QY 182 sTyrlleGluThrProLeuLeuIleCysAspThrLysPheAspIleArgGlnTrpPhe 202
|||
|||
Db 811 ATATATCTCCAACTCTTACTTATTGGCAGATATAAATGTGATCTCCGATCTTATGTTTG 870
|||
|||
QY 202 uValThrAspTrpAsnProLeuThrIleTrpPheTyrLysGluSerTyrLeuArgPhe 222
|||
|||
Db 871 TGTACTGCTTAAAGCTTTGACCAATTTATGTTTATCAGGAAGGTTGGTTCGGTTTCG 930
|||
|||
QY 222 rThrGlnArgPheSerLeuAspLysLeu---AspSerAlaIleHisLeuCysAsnAl 241
|||
|||
Db 931 CACGGAAGAGTTGACCTCAGTAATTTGCCAAACAATATTATGCCCATTTGACCAACAGCAG 990
|||
|||
QY 241 aValGlnLys-----TyrLeuLysAsnAspValGlyArgSerProLe 255
|||
|||
Db 991 CATCAATAAATCCGGGGCTCTTATGAGAAGATCAAGAAGTGTGGTCATGGT----- 1045
|||
|||
QY 255 uLeuProAlaHisAsnMetTrpThrArgPheGlnGluTyrLeuGlnArgGlnG 275
|||
|||
Db 1046 -----TGTAAATGGACGCTCAGCAGATTTTCTTACCTTCGTAGC----- 1087
|||
|||
QY 275 yArgGlyAlaValTrp-----GlySerValIleTyrProSerMetLysLysAlaI 292
|||
|||
Db 1088 -----TGGGATGTGGACGATCTGCTTTTGTGGAAGAAATCCACCGCATGCT 1134
|||
|||
QY 292 eAlaHisAlaMetLysValAlaGlnAspHisValGluProArgLysAsnSerPheGlu 312
|||
|||
Db 1135 TATCTTCACCATCTCGCATTCGCCATCTCTCCCTTTTGTGCCAATTCGTTTGGAGCT 1194
|||
|||
QY 312 uTyrGlyAlaAspPheValLeuGlyArgAspPheArgProTrpLeuIleGluIleAsnSe 332
|||
|||
Db 1195 CTTTGGGTTTGATATTTTGTATTTGATGACNACTTGAACCATTCGCTTTTGTAGGTCACTA 1254
|||
|||
QY 332 rSerPro-----ThrMetHisProSerThrProVal 342
|||
|||
Db 1255 CAGCCCGACCTTGACCTTGATGTTGTTCAACAGATGTG 1291
|||
|||
```

RESULT 5

```
US-09-270-767-15138
; Sequence 15138, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: 7326-094
```

```
; CURRENT APPLICATION NUMBER: US/09/270, 767
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15138
; LENGTH: 1980
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-15138
```

Alignment Scores:

```
Pred. No.: 4,718-11 Length: 1980
Score: 228.50 Matches: 83
Percent Similarity: 40.7% Conservatives: 53
Best Local Similarity: 24.9% Mismatches: 110
Query Match: 8.0% Indels: 88
DB: 3 Gaps: 11
```

US-10-635-977-2 (1-541) x US-09-270-767-15138 (1-1980)

```
QY 179 ValValGlnLysTyrIleGluThrProLeuLeuIleCysAspThrLysPheAspIleArg 198
|||
|||
Db 1 GTGGTGTGCAATATATAGTGGATCGCTGTGCTGATGTCACAAATGTGATTGAGG 60
|||
|||
QY 199 GlnTrpPheLeuValThrAspTrpAsnProLeuThrIleTrpPheTyrLysGluSerTyr 218
|||
|||
Db 61 GTATACGTTCTAGTCACCTTCCTTCGATCCCTTAATATTATTATATGAGAGGCAAT 120
|||
|||
QY 219 LeuArgPheSerThrGlnArgPheSer-----LeuAspLysLeu---AspSerAlaIle 235
|||
|||
Db 121 GTGAGATTGGCCACAGTCAAGTACGACGAGGATCCGATAACCTGTGGAAACCCGTGCATG 180
|||
|||
QY 236 HisLeuCysAsnAsnAlaValGlnLysTyrLeuLysAsn----- 248
|||
|||
Db 181 CATCTCTGCAATTACAGCATTAACAAATACCATCTGATTACATAGGAGCTCCGATGCT 240
|||
|||
QY 249 -----AspValGlyArgSerProLeuLeuProAlaHisAsnMetTrpThrSerThr 265
|||
|||
Db 241 CAGGATGAGGATGTAGGC-----CATAAGTGGACGTTGTGTC 276
|||
|||
QY 266 ArgPheGlnGluTyrLeuGlnArgGlnGly-----ArgGlyAlaValTrpGly--- 281
|||
|||
Db 277 GCTCTTTTAAAGCACTGAAACTCCAGAGTGTGTGATACACGTCAGCTAATGCTGAATAT 336
|||
|||
QY 282 ---SerValIleTyrProSerMetLysLysAlaIleAlaHisAlaMetLysValAlaGln 300
|||
|||
Db 337 GAGGATTTAATTTAAGCTGTGTAGCTGTGCTCAATCTATTATTTCGGCCTGCAGG 396
|||
|||
QY 301 AspHisValGluProArgLysAsnSerPheGluLeuTyrGlyAlaAspPheValLeuGly 320
|||
|||
Db 397 ATGTTCTGTTCCCAATGGAATAAATCTGTTGAGCTGTATGGATTGATATATAATAGAC 456
|||
|||
QY 321 ArgAspPheArgProTrpLeuIleGluIleAsnSerSerProThrMetHisProSerThr 340
|||
|||
Db 457 AACGCACATAAGCCATGCTGCTGAAATCAATCTATCCGCTCAATGGGTGGTGGATAGT 516
|||
|||
QY 341 ProValThrAlaGlnLeuCysAlaGlnValGlnGluAspThrIleLysValAlaValaAsp 360
|||
|||
Db 517 CTTTGGACACTAAGTT----- 534
|||
|||
QY 361 ArgSerCysAspIleGlyAsn-PheGluLeuLeuTrpArg---GlnProValValGluPr 379
|||
|||
Db 535 AAGTCATGCTGATGGCAGACCTATTGACGTTGTGGGCATTTCCAGCCTATAGTC----- 589
|||
|||
QY 379 oProProPheSerGlySerAspLeuCysValAlaGlyValSerValArgArgAlaArgAr 399
|||
|||
Db 589 ----- 589
|||
|||
QY 399 gGlnValLeuProValCysAsnLeuLysAlaSerAlaSerLeuLeuAspAlaGlnProle 419
|||
|||
Db 590 -----CCGAATGAATCTCATTTATGACCAAGAAATGGTCGCGTTCCGAGCTCA 639
|||
|||
QY 419 uLysAlaArgGlyProSerAlaMetProAspProAlaGlnGlyProProSerProAlaIle 439
|||
|||
```

```

Db      640 AGTGCAGAAAGGATAGCAACTTTTCCAGTACCTCCAAAAGCAAGAAGT----- 691
Qy      439 uGlnArgAspLeuGlyLeuGlyGluGlyGlyLeuProLeuAlaLeuLeuAlaProLe 459
Db      692 -----CGAAGAAAAAAGGAGCTGCCATTAACTTAATCTTAATCTAAACCGGC 732
Qy      459 uArgGlyAlaAla-----GluSerGlyG1 467
Db      733 GAAGAGCAGCGCATATTGCGCAATGCAAGCTACAGTATTCTCGTGTGAGGATTCGTG 792
Qy      467 yAlaAlaGlnProThrArgThrLysAlaAlaGlyLysVal 480
Db      793 CGCATCTTTCCACAGATGACTCAATGCACGCGTATGGTA 832

```

RESULT 6

```

US-10-104-047-1651
; Sequence 1651, Application US/10104047
; Patent No. 6943241
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. 6943241el full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; PRIOR FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1651
; LENGTH: 2111
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-104-047-1651

```

Alignment Scores:

```

Pred. No.:      2,66e-09      Length:      2111
Score:          210.00      Matches:      146
Percent Similarity: 32.7%      Conservative: 53
Best Local Similarity: 24.0%      Mismatches: 202
Query Match:      7.4%      Indels:      210
DB:              3      Gaps:      31

```

US-10-635-977-2 (1-541) x US-10-104-047-1651 (1-2111)

```

Qy      6 LeuLysTrpValValSerHisGlnSerCysSerArgSerSerArgSerLysProArgAsp 25
Db      532 CTGAAGTGG-----TGTGAGTCAAGAGCCGAGACAGCTACGGCAGC 573
Qy      26 GlnArgGluGluAlaGlySerSerAspLeuSerSerArgGlnAspAla----- 41
Db      574 TTCGGGAAGG-AGAGCAGCTGCTGTACACAGCTTCCCAACAACAAGCTCTCACCACCAA 632
Qy      42 GluAsnAlaGlu-AlalysLeuArgGlyLeuProGlyGlnLeuValAspIleAlaCysLY 61
Db      633 GATCGGCGTCTCAGCACCCCTTCGGGA---CGGGCACCGGCCCATGAGCGGGCCACAA 689
Qy      61 sValCysGlnAlaTyrlLeuGlyGlnLeuGluHisGluAspIleAspThrSerAlaAspAl 81
Db      690 GGTGCCGGGGGGTCCAGGCCAGCTGGAAAG-----GACGCAGCAGCGCCCGC 740
Qy      81 aValGluAspLeuThrGluAlaGluTrpGluAspLeuThrGlnGlnTyrlSerLeuVa 101
Db      741 CCTGGAGGACCTC-----CCGTGG----- 759
Qy      101 lHisGlyAspAlaPheIleSerAsnSerArgAsnTyrlPheSerGlnCysGlnAlaLeuLe 121
Db      760 -----ACAAGCCCGAGATACCTCAGGCCACAGAGGGTCCGT-- 795
Qy      121 uAsnArgIleThrSerValAsnProGlnThr-----AspIleAspGlyLeuArg-- 137
Db      796 ----AGATGGAAGAGTTTTTCCAGAGACCTTACCGCTGGACCTCAACACAGAGAGAGA 851

```

```

Qy      138 -----AsnIleTrpIleIleLysProAlaAlaLysSe 148
Db      852 GGCCTTTTTCACCTGTTTGAATGAACCCAGATATGGATCTGCAGCCACAGCCTCCAA 911
Qy      148 zArgGlyArgAspIleValCysMetAspArgValGluGluIleLeuGluLeuAlaAla-- 167
Db      912 CCAGGGCAAAAGCATCTTCTGCTCGGAACACAGAGGAAGTTCGCCCTCGAGCCAA 971
Qy      168 -----AlaAspHisProLeuSerArgAspAsnLysTrp----- 178
Db      972 GACCCGAGCATGGAGGACGACCCCATCCACCAAGACGCCGCTCCGGGGGCTCAGGC 1031
Qy      179 ----ValValGlnLysTyrlleGluThrProLeuLeuIleCysAspThrLysPheAspIl 197
Db      1032 GCGGTGTGTCAGAGGTATTCAGAACCCGCTGCTGTGGACGGAGAAAGTTTGACGT 1091
Qy      197 eArgGlnTrpPheLeuValThrAspTrpAsnProLeuThrIleTrpPheTyrlLysGluSe 217
Db      1092 GCGCTCTACCTGCTCATTTGCTGTGCACACACCCATCATGATC---TTCTTTGGCCACCG 1148
Qy      217 rTyrlLeuArgPheSerThrGlnArgPheSerLeuAspLysLeuAspSerAlaIleHisLe 237
Db      1149 CTATGCTCGCTCACCCCTTAGCCTTTACGACCCCATTCACGACGCTCGCGGCCCAT 1208
Qy      237 uCysAsnAsnAlaValGlnLysTyrlLeuLysAsnAspValGlyArgSerPro----- 254
Db      1209 GACCAACCAAGTTTCATCAGAAG-----AAGAGCCTCTGTACAT 1247
Qy      255 -LeuLeuProAlaHisAsnMetTrpThrSerThrArgPheGlnGlyLeuGlnArgG1 274
Db      1248 GCTGCTGAAGGAGCACACGCTGTGGAGC-----ATGGAACATCTCAACCGCTA 1295
Qy      274 nGlyArgGlyAlaValTrpGlySerValIleTyrlProSerMetLysAlaIleAlaHi 294
Db      1296 CATCAGTGACACGTTCTGG-----AA 1316
Qy      294 sAlaMetLysValAlaGlnAspHisValGluProArgLysAsnSerPheGluLeuTyrlG1 314
Db      1317 GGCCTGGGGCTCGCCCAAGGAC----- 1338
Qy      314 yAlaAspPheValLeuGlyArgAspPheArgProTrpLeuIleGluIleAsnSerSerPr 334
Db      1339 -----TGGGTCTTC----- 1347
Qy      334 oThrMetHisProSerThrProValThrAlaGlnLeuCysAlaGlnValGlnGluAspTh 354
Db      1348 -----ACCACCTCAAGGTGCGTCCACTGTGC----- 1374
Qy      354 rIleLysValAlaValAspArgSerCysAspIleGlyAsnPheGluLeuLeuTrpArgG1 374
Db      1375 -----CC 1376
Qy      374 nProValValGluProProPheSerGlySerAspLeuCysValAlaGlyValSerVa 394
Db      1377 TCCAGTCTGGAGTGTAGATCCC---TCGGGGGGGGG---TGTGTGTGAGGTGGGCAC 1430
Qy      394 lArg-----ArgAlaArgArgGlnValLeuProValCy 405
Db      1431 CAGGCACACAGATGTCCGTGCGTGGCGGCTGCTGCTGAGTGTGACCTGACCGCTG 1490
Qy      405 sAsnLeuLys-----AlaSerAlaSerLeuLeuAspAlaGlnProLeuLy 420
Db      1491 TGAACCAACCCCTTCCAGCGTCTGTCTCATTTAGCTGGCAGTGCCTGCTCCAGCAGC 1550
Qy      420 sAlaArgGlyProSerAlaMetProAspProAlaGlnGlyProProSerProAlaLeuG1 440
Db      1551 CAGCAAGGCCCGAGCGCAAGCCAGTCCGGGTCTGTCTGGCAGCAGTCCCGGGGCGAG 1610
Qy      440 nArgAspLeuGlyLeuLysGluGlyGlyLeuProLeuAla-----LeuLe 456
Db      1611 CCTCGCGTAGGGTCAGGGGACAGCTCGGCTTCAGTGACAGCCACCATGTGAAGAGCTTT 1670
Qy      456 uAlaProLeuArgGlyAlaAlaGluSerGlyGlyAlaAlaGlnProThrArgThrLysAl 476

```

Db 1671 TGTTCCTTCAGATCTCCACCTGGCAGGAGG-----CCCCCAGCAGCCCCCGG 1721
Qy 476 aLaGlyLysValGlu-LeuProAlaCysProCysArg----- 488
Db 1722 GCCAGGTGCTGGAGCAGCAGCAGCTGCCGTGCGAGGCCGCGACTCTCCCTAGCTAAGG 1781
Qy 489 --HisValAspSerGlnAlaProAsn-----ThrGlyValProVala 502
Db 1782 GCCATGTGCG-GTGGAGTGTCTAAATTTTGCAGGGTTCCTTCTGTGGGGGTACCTGTGG 1840
Qy 502 la-----GlnProAlaLysSer-----TTPAspP 510
Db 1841 GGTACTTCAACAGCCCTAGCAGCAAGGCCCTTGAGCAGCGCGGTGTCAAACTGGGATA 1900
Qy 510 roAsnGlnLeuAsnAlaHisProLeuGluProValLeuArgGlyLeuLysThrAlaGluG 530
Db 1901 AAAAC-----GGGGCTTGGCCGAGGACTTTTATCTG-----TCIT 1936
Qy 530 lyAlaLeuArgProProGly 537
Db 1937 GGTACCCCGCCGAGCCCGGA 1959

RESULT 7

US-09-799-451-295
; Sequence 295, Application US/09799451
; Patent No. 6783969
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Zhou, Ping
; APPLICANT: Goodrich, Ryle
; APPLICANT: Asundi, Vinod
; APPLICANT: Ren, Feiyan
; APPLICANT: Zhang, Jie
; APPLICANT: Xue, Aidong J.
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Ma, Yunqing
; APPLICANT: Yamazaki, Victoria
; APPLICANT: Chen, Rui-hong
; APPLICANT: Wang, Zhiwei
; APPLICANT: Wang, Dunrui
; APPLICANT: Yang, Yonghong
; APPLICANT: Wehrman, Tom
; APPLICANT: Ghosh, Reena
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. 6783969el Nucleic Acids and
; FILE REFERENCE: Polypeptides
; CURRENT APPLICATION NUMBER: US/09/799,451
; CURRENT FILING DATE: 2001-03-05
; NUMBER OF SEQ ID NOS: 948
; SOFTWARE: pt_FL_genes Version 2.0
; SEQ ID NO 295
; LENGTH: 2716
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (277) .. (2442)
US-09-799-451-295

Alignment Scores:
Pred. No.: 1,02e-08 Length: 2716
Score: 205.50 Matches: 78
Percent Similarity: 42.2% Conservative: 49
Best Local Similarity: 25.9% Mismatches: 116
Query Match: 7.2% Indels: 59
DB: 3 Gaps: 11

US-10-635-977-2 (1-541) x US-09-799-451-295 (1-2716)

Qy 106 PheIleSerAsnSerArgAsnTyrPheSerGlnCys----- 117

Db 80 TATGCTAGTTTGAATAGAAATTAATTTTCTTATTATGTCCAATTTTACAGAAGGATCAAGT 139
Qy 118 -----GlnAlaLeuLeuAsnArgIle 124
Db 140 CTCGGCCTCTGGATTATACCTTTTGTTCCTCGAAGCTTGGATCTTCTCTGCTGAATATACTC 199
Qy 125 ThrSerValAsnProGlnThrAspIleAspGlyLeuArg-----AsnIleTrpIleIle 142
Db 200 AATTCCAA-AATTATGTGAAGAAATTGAAGAAAAACCGAAGCAGAAAACTTTTATAGTG 258
Qy 143 LysProAlaAlaLysSerArgGlyArgAspIleValCysMetAspArgValGluGluIle 162
Db 259 AAACGAGCTAATGGTGCAATGGGTCTATGGGATTTCTTGTATAGAAAAATGGTGAC----- 312
Qy 163 LeuGluLeuAlaAlaAspHisProLeuSerArgAspAsnLysTrpValValGlnLys 182
Db 313 ---AACTTCCATCTCAGATCAT-----TTGATTGTTCAAGAA 348
Qy 183 TyrIleGluThrProLeuLeuIleCysAspThrLysPheAspIleArgGlnTrpPheLeu 202
Db 349 TACATTGAAAAGCCTTCTTAATGGAAGGTTACAAGTTTGACTTACGAATTTATATTCTG 408
Qy 203 ValThrAspTrpAsnProLeuThrIleTrpPheTyrLysGluSerTyrLeuArgPheSer 222
Db 409 GTTACATCGTGTGATCCACTAAAAATATTTCTCTACCATGATGGGCTTGTGCGAATGGGT 468
Qy 223 ThrGlnArgPhe-----SerLeuAspLysLeuAspSerAlaIleHis 236
Db 469 ACAGAGAAAGTATCTCCACCTAATGAGTCCCAATTTGACCCAGTTA-----TACATGCAT 522
Qy 237 LeuCysAsnAsnAlaValGlnLys-----TyrLeuLysAsnAspValGlyArgSer 253
Db 523 CTGACAAACTACTCCGTGAACAAGCATAATGAGCATTTTGAACGGGATGAAACTGAGAAC 582
Qy 254 ProLeuLeuProAlaHisAsnMetTrpThrSerThrArgPheGlnGluTyrLeuGlnArg 273
Db 583 AAAGGCAGCAAAACGTTCCATCAAAATGG-----TTTACAGAAATTCCTTCAAGCA 630
Qy 274 GlnGlyArgGlyAlaVal-----TrpGlySerValIleTyrProSerMetLysLysAla 291
Db 631 AATCAACATGATGTTGCTAAGTTTTGGAGTGATATTTTCCAGAAATGGTGTGAAGACCCTG 690
Qy 292 IleAla-----HisAlaMetLysValAlaGlnAspHisValGluPro 305
Db 691 ATTGTAGCAGAACCTCATGTCTCTGCATCGCTATCGAATGTGTAGACCTGGTCAACCTCCA 750
Qy 306 ArgLysAsnSer-----PheGluLeuTyrGlyAlaAspPheValLeuGlyArgAspPhe 323
Db 751 GGAAGCGAAAGTGTCTGCTTTGAAAGTCTCGGATTTTGATATTTTGTGATAGAAAACTA 810
Qy 324 ArgProTrpLeuIleGluLeuAsnSerSerProThrMetHisProSerThrProValThr 343
Db 811 AAGCCATGGCTTCTGGAGATTAAACCGACCCCAAGCTTTTGGAACTGTAGCAAAAATAGAC 870
Qy 344 AlaGlnLeuCysAlaGlnValGlnGluAspThrIleLysValAlaValAspArgSerCys 363
Db 871 TATGATGTAAAAAGGGGAGTGCTGTCTAAATCGTTGAAGCTTACTATAAAGGACCAGT 930
Qy 364 Asp 364
Db 931 GAC 933

RESULT 8

US-09-814-915A-56
; Sequence 56, Application US/09814915A
; Patent No. 6750015
; GENERAL INFORMATION:
; APPLICANT: Horwitz, Kathryn
; APPLICANT: Richer, Jennifer
; TITLE OF INVENTION: Progesterone Receptor-Regulated Gene Expression and Methods Rel
; TITLE OF INVENTION: Theredo
; FILE REFERENCE: 2848-39

; CURRENT APPLICATION NUMBER: US/09/814,915A
; CURRENT FILING DATE: 2002-03-21
; PRIOR APPLICATION NUMBER: 60/214,970
; PRIOR FILING DATE: 2000-06-28
; NUMBER OF SEQ ID NOS: 108
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 56
; LENGTH: 3394
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-814-915A-56

Alignment Scores:
Pred. No.: 6,43e-05 Length: 3394
Score: 166.00 Matches: 132
Percent Similarity: 32.6% Conservative: 64
Best Local Similarity: 22.0% Mismatches: 206
Query Match: 5.8% Indels: 202
DB: 3 Gaps: 25

US-10-635-977-2 (1-541) x US-09-814-915A-56 (1-3394)

```
QY      8  TipValValSerHisGlnSerCysSerArgSerSerArgSer----- 21
DB      874 TGGTGCACCCCGCCGACATCTTCAGGCTTACAGGAGCTGCAGCGTGCAGCGTGCAGCA 933

QY      22 -----LysProArgAspGlnArgGluGluAlaGlySerSer 33
DB      934 GCCTCACCCACCGCGCTTCCACCTCACCGAGTGAGCGGCGCGCATCTCTTCA 993

QY      34 AspleuSerSerArgGlnAspAlaGluAsnAlaGluAlaLysLeuArgGlyLeuProGly 53
DB      994 ACTTCTCACACTTCAAGGACTACAGGAACCTCA-----GCCAGGAG-AGGCCAGGC 1043

QY      54 GlnLeuValAspIleAlaCysLysValCysGlnAlaTyrLeuGlyGlnLeuGluHisGlu 73
DB      1044 GTGCTG-----CTGAACCAAGTTCCCTCCGCGAG 1070

QY      74 AspleuAspThrSerAlaAspAlaValGluAspLeuThrGluAlaGluTrpGluAspLeu 93
DB      1071 AACCTGCTGACTGTCAAGGACTGCTGCGCTCCATCGCGCGCGGCGAGGTGGCGCC-- 1127

QY      94 ThrGlnGlnTyrTyrSerLeuValHisGlyAspAlaPheIleSerAsnSerArgAsnTyr 113
DB      1128 -----GAGGGCCCAACCTGGTGGTGGCCCGAACCTTCAACCTG 1163

QY      114 PheSerGlnCysGlnAlaLeuLeuAsnArgIleThrSerValAsnProGlnThrAspIle 133
DB      1164 CGCACTGAG-----CTGCCCCAGTTTGTCTAGCTACTTCCAGCAGCGGGAAGG 1211

QY      134 AspGlyLeuArgAsnIleTrpIleLysProAlaAlaLysSerArgGlyArgAspIle 153
DB      1212 TGGGGCGGAGACAACCACTGGATCTGCAAGCCCTGGAACCTGGCGCGCAGCTGGAC-- 1268

QY      154 ValCysMetAspArgValGluGluIleLeuGluLeuAlaAlaAspHisProLeuSer 173
DB      1269 -----ACCCAGCTCACCAAGAG-CCTGCACAGCATCATCCGGCACCG 1309

QY      174 ArgAspAsnLysTrp-ValValGlnLysTyrIleGluThrProLeuLeu----- 189
DB      1310 AGAGCAGACCCCAAGGTGTGTTCAGTATCATCGAAGTCCCGTGTGTCTTCGAGA 1369

QY      190 ----IleCysAspThrLysPheAspIleArgGlnTrpPheLeuValThrAspTrpAsnPr 208
DB      1370 AGACGTGGGAAGGTCAAGTTCGACATCCGCTACATCGTGTCTGGGTCAAGTAGGCC 1429

QY      208 oLeuThrIleTrpPheTyrLysGluSerTyrLeuArgPheSerThrGlnArgPheSerLe 228
DB      1430 CCTACGTTGTTCGTGTATGATGTGTTCTGGCTCGGTTCTTCAACCGGGCTTTGCACT 1489

QY      228 uAspLysLeuAsp----- 232
DB      1490 CAACGACCTGGATGACTACGAGAGCACTTCCAGCGTCAATGAACCTATGACCCCGATGGT 1549
```

RESULT 9

US-09-270-767-31557
; Sequence 31557, Application US/09270767
; Patent No. 6703491

```
QY      233 ----SerAlaIleHisLeuCysAsnAsnAlaValGlnLysTyrLeuLysAsnAspValG 251
DB      1550 GCTGAAGCAGGTGCAC---TGTGAAGAGTTATCCCGAGTTTGAGAG----- 1595

QY      251 yArgSerProLeuLeuProAlaHisAsnMetTrpThrSerThrArg----- 266
DB      1596 -CAATACCCAGAAATTTCCC-----TGGACGGACGTCACGGCTGAGATCTTCG 1642

QY      267 ----PheGlnGluTyrLeuGln---ArgGlnGlyArgGlyAlaValTrpGlySerVal 284
DB      1643 GGCCTTTCAGGAGCTGTTCAGGTGGCTGTGTCAGACCCACCACCCCTGGGCTCTGCG 1702

QY      284 eTyrProSerMetLysLysAlaIleAlaHisAlaMetLysValAlaGlnAspHisValG 304
DB      1703 CTACCCCTCA----- 1712

QY      304 uProArgLysAsnSerPheGluLeuTyrGlyAlaAspPheValLeuGlyArgAsp----- 322
DB      1713 -----TCCCGGGCCATGTATGCCGTGACCTCATGCTGAAGTGGGACAACGG 1759

QY      323 -----PheArgProTrpLeuIleGluLeuAsn-SerSerProThrM 336
DB      1760 CCCAGATGGAAGCGGCTGATGCAGCCGAGATCTCTGGAGGTGAATTCACCCCGCAGCTG 1819

QY      336 etHis-----ProSerThrProValThrAlaGlnLeuCysAlaG 349
DB      1820 TGAGCAGGCTGAGTACACCCACCTTCTTCAAGCAGCTTTCAGACCTTGT----- 1875

QY      349 lnValGlnGluAspThrIleLysValAlaValAspArgSerCysAspIleGlyAsnPheG 369
DB      1875 ----- 1875

QY      369 luLeuLeuTrpArgGlnProValValGluProProPheSerGlySerAspLeuCysV 389
DB      1876 ----TTCGACCGCGGCTGCTGCCACGTACCT----- 1908

QY      389 alAlaGlyValSerValArgAlaArgGlnValLeuProValCysAsnLeuLysA 409
DB      1909 -----GCCTGTCTTAGGCACTCGCTGTCCCAAACTGTGTGGGGC-AGGA 1956

QY      409 laSerAlaSerLeuLeuAspAlaGlnProLeuLysAlaArgGlyProSerAlaMetProA 429
DB      1957 TTCAACCTCAGTTCTCTGAGCTGCTTCTGCAAGGCC-----CCCATGTCTCTCCCA 2010

QY      429 spProAlaGlnGly-----ProProSerProAlaLeuGlnA 441
DB      2011 CACCGGCTCTGGGCTAGAGCTCAGCCCGCAGGCTCTGTCTCTGCGAGCCATCTCCCGGC 2070

QY      441 rgAspLeuGlyLeuLysGluLysGlyLeuProLeuAlaLeuLeuAlaProLeuArg 461
DB      2071 GCCAC-----ACTCGGAGGACACAGCATCTCTCTCTACCT 2106

QY      461 lyAlaAlaGluSerGlyGlyAlaAlaGlnProThrArgThrLysAlaAlaGlyLysValG 481
DB      2107 GTGGTCAGAGCAGGACAGTGTGTTGTCCTCC-----AGGCTG 2145

QY      481 luLeuProAlaCysProCysArgHisValAspSerGlnAlaProAsnThrGlyValProV 501
DB      2146 AGCACCACCCCGCCCTGCCTCAC-----CCCTCACCCCATCTGTGCACCTGATGAG 2199

QY      501 alAlaGlnProAlaLysSerTrpAspProAsnGlnLeuAsnAlaHisProLeuGluProV 521
DB      2200 TCTCAGTTAGCAAGGGCTTGTCTCTGGCATGGGAATTTGTTTC---CTGGCTGCTG 2256

QY      521 alLeuArgGlyLeuLysThrAlaGluGlyAlaLeuArgProProGlyGlyLysGly 540
DB      2257 TGTT-----CCAGGGGCTGCTGGG 2276
```

; GENERAL INFORMATION:
 ; APPLICANT: Homburger et al.
 ; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
 ; FILE REFERENCE: File Reference: 7326-094
 ; CURRENT APPLICATION NUMBER: US/09/270,767
 ; CURRENT FILING DATE: 1999-03-17
 ; NUMBER OF SEQ ID NOS: 62517
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 31557
 ; LENGTH: 783
 ; TYPE: DNA
 ; ORGANISM: Drosophila melanogaster
 US-09-270-767-31557

Alignment Scores:
 Pred. No.: 7,59e-05 Length: 783
 Score: 154.50 Matches: 40
 Percent Similarity: 46.0% Conservative: 18
 Best Local Similarity: 31.7% Mismatches: 43
 Query Match: 5.4% Indels: 25
 DB: 3 Gaps: 3

US-10-635-977-2 (1-541) x US-09-270-767-31557 (1-783)

QY 125 ThrSerValAsnProGlnThrAspIle----- 133
 Db 406 ACCACCTATAATCTGTCCACCGAAGTAAAGGAGTTTCGCGGCTTACTACCAACGAGAGCA 465
 QY 134 ---AspGlyLeuArgAsnIleTPIlelleLysProAlaLysSerArgGlyArgAsp 152
 Db 466 GCCNAGGGCTCGACAAACCACTGGATTATCAAGCCCTGGAACCTTGGCTCGCGATTGGAC 525
 QY 153 IleValCysMetAspArgValGluGluIleLeuGluLeuAlaAlaAspHisProLeu 172
 Db 526 ACACATATCAGCGACACATCAAGCAGATCGTAGCTACCGGCCACT----- 573
 QY 173 SerArgAspAsnLysTrpValValGlnLysTyrIleGluThrProLeuLeuIleCys--- 191
 Db 574 -----GGACCGAAGATAGCGCAGAAGTACATCGAGCGACCACTTCTTTAGTCGG 624
 QY 192 -----AspThrLysPheAspIleArgGlnTrpPheLeuValThrAspTrp 206
 Db 625 CAGGAGGTGGAAGCGAGCGTTAAGTTCGATATTCGCTATGTAATCTGTGTGAAAAGCGTC 684
 QY 207 AsnProLeuThrIleTrpPheTyrLysGluSerTyrLeuArgPheSerThrGlnArgPhe 226
 Db 685 AAACCCCTCAGGCATATATACACCGAAGTCTTCTTCGCTTTCGCAAAACCATCCCTTC 744
 QY 227 SerLeuAspLysLeuAsp 232
 Db 745 ACGCTGGATCACTTCGAT 762

RESULT 10
 US-09-270-767-15222
 ; Sequence 15222, Application US/09270767
 ; Patent No. 6703491
 ; GENERAL INFORMATION:
 ; APPLICANT: Homburger et al.
 ; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
 ; FILE REFERENCE: File Reference: 7326-094
 ; CURRENT APPLICATION NUMBER: US/09/270,767
 ; CURRENT FILING DATE: 1999-03-17
 ; NUMBER OF SEQ ID NOS: 62517
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 15222
 ; LENGTH: 1525
 ; TYPE: DNA
 ; ORGANISM: Drosophila melanogaster
 US-09-270-767-15222

Alignment Scores:
 Pred. No.: 0.000214 Length: 1525
 Score: 154.50 Matches: 40

Percent Similarity: 46.0% Conservative: 18
 Best Local Similarity: 31.7% Mismatches: 43
 Query Match: 5.4% Indels: 25
 DB: 3 Gaps: 3

US-10-635-977-2 (1-541) x US-09-270-767-15222 (1-1525)

QY 125 ThrSerValAsnProGlnThrAspIle----- 133
 Db 1148 ACCACCTATAATCTGTCCACCGAAGTAAAGGAGTTTCGCGGCTTACTACCAACGAGAGCA 1207
 QY 134 ---AspGlyLeuArgAsnIleTPIlelleLysProAlaLysSerArgGlyArgAsp 152
 Db 1208 GCCAAGGGCTCGACAAACCACTGGATTATCAAGCCCTGGAACCTTGGCTCGCGATTGGAC 1267
 QY 153 IleValCysMetAspArgValGluGluIleLeuGluLeuAlaAlaAspHisProLeu 172
 Db 1268 ACACATATCAGCGACACATCAAGCAGATCGTAGCTACCGGCCACT----- 1315
 QY 173 SerArgAspAsnLysTrpValValGlnLysTyrIleGluThrProLeuLeuIleCys--- 191
 Db 1316 -----GGACCGAAGATAGCGCAGAAGTACATCGAGCGACCACTTCTTTAGTCGG 1366
 QY 192 -----AspThrLysPheAspIleArgGlnTrpPheLeuValThrAspTrp 206
 Db 1367 CAGGAGGTGGAAGCGAGCGTTAAGTTCGATATTCGCTATGTAATCTGTGAAAAGCGTC 1426
 QY 207 AsnProLeuThrIleTrpPheTyrLysGluSerTyrLeuArgPheSerThrGlnArgPhe 226
 Db 1427 AAACCCCTCAGGCATATATACACCGAAGTCTTCTTCGCTTTCGCAAAACCATCCCTTC 1486
 QY 227 SerLeuAspLysLeuAsp 232
 Db 1487 ACGCTGGATCACTTCGAT 1504

RESULT 11
 US-09-902-540-1249
 ; Sequence 1249, Application US/09902540
 ; Patent No. 6833447
 ; GENERAL INFORMATION:
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Wiegand, Roger C.
 ; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
 ; FILE REFERENCE: 38-10(15849)B
 ; CURRENT APPLICATION NUMBER: US/09/902,540
 ; CURRENT FILING DATE: 2001-07-10
 ; PRIOR APPLICATION NUMBER: 60/217,883
 ; PRIOR FILING DATE: 2000-07-10
 ; NUMBER OF SEQ ID NOS: 16825
 ; SEQ ID NO 1249
 ; LENGTH: 30135
 ; TYPE: DNA
 ; ORGANISM: Myxococcus xanthus
 US-09-902-540-1249

Alignment Scores:
 Pred. No.: 0.0304 Length: 30135
 Score: 153.00 Matches: 160
 Percent Similarity: 33.8% Conservative: 96
 Best Local Similarity: 21.1% Mismatches: 240
 Query Match: 5.4% Indels: 263
 DB: 3 Gaps: 3

US-10-635-977-2 (1-541) x US-09-902-540-1249 (1-30135)

QY 14 SerCysSerArgSerSerArgSerLysProArgAspGlnArgGluGluAlaGlySerSer 33
 Db 7669 TCCACGCGAGCGGTGACGAAGACGCCCGCGAGACCTTCGCGCACGACGCCCGGG 7728
 QY 34 AspLeuSerSer-----ArgGln-----AspAlaGluAsnAlaGluAlaLys 47


```
Db 7729 GAC TTCGGGACCCCGAGCGCCCTTCGCGAGGAGATTCTCCCTCCGCTGGGCC 7788
Qy 48 LeuArgly---LeuProGlyGlnLeu-ValAspIleAlaCysLys---ValCysGlnAl 65
Db 7789 GCTCAGGATGAGCGCTGCGCGCTTCGATTGACAGGCTTTCGAGAGGTCTTCCAAG 7848
Qy 65 a-----TyrLeuGlyGlnLeuGl 71
Db 7849 CGAGCGTGGCGAGCGGGCCGAGCGCTCCAAAGTGGATGGACTTCGACGGAGCTCCG 7908
Qy 71 uHisGluAspIleAspThrSerAlaAsp-----AlaValGluAspLeuThrGl 87
Db 7909 CACACAGAACCCGGAATACACCGCGGAGGAGTCCGCTTCGCCATCCAGACAAAGTGGC 7968
Qy 87 uAlaGluTrpGluAspLeuThrGlnGlnTyrTrpSerLeuValHisGlyAspAlaPheIl 107
Db 7969 ATTGGAGCGGACCGACTG-----GAGCGCCATTCCGCGCGAACAACACTGACAGGTTCAT 8022
Qy 107 eSerAsnSerArgAsnTyrPheSerGlnCys----- 117
Db 8023 CTCGAGGCCATCGCGTGGGTCAACCACTGCTACGACAAAGCGCCCTCCACGGA 8082
Qy 118 -----GlnAlaLeuLeuAsnArgIleThrSerValAsnProGlnTh 131
Db 8083 GCGCGACATCAGCAGCTGGCGGCTTCGCGCGGACCTGATGAGCAGATCCAAAGAC 8142
Qy 131 rAspIleAspGlyLeuArgAsnIleTrpIleIleLysProAlaAlaLysSerArgGlyAr 151
Db 8143 GACCCGAGGAGCTCAAGTACGCATCATCGAGCGCTCGCGGAGGTGACGGGAAC 8202
Qy 151 gAspIleValCysMetAspArgValGluGluIleLeuGluLeuAlaAlaAspHisPr 171
Db 8203 GGACGCGGTGACGCGCGAGAACATCGAAGTTTCATCCAGGACCGCTC----- 8251
Qy 171 oLeuSerArgAspAsnLysTrpValValGlnLysTyrIleGlu-----Th 186
Db 8252 -----AAGTGGTTCACCCAGTGTCTACGAGCAGACGAGGCGCGAGGCCAC 8295
Qy 186 rProLeuLeuIleCysAspThrLysPheAspIleArgGlnTrp----- 200
Db 8296 GCCCGC-----GAGTGAAGGAGTGGCGCGCTTCGCCCAGGA 8334
Qy 201 -----PheLeuValThrAsp-- 205
Db 8335 GCATTTGAAGAAGACCCGGAATCTCCCGGAGCACCTCAAGTACGCCATCCCGACG 8394
Qy 206 -----TrpAsnProLeuThrIleTrpPheTyrLy 215
Db 8395 CATTCGCGGAGATGACGGGACCGGTGAGCGACCGCTGACATCGACAGTTTCAT 8454
Qy 215 sGluSerTyrLeuArgPheSerThrGlnArgPhe---SerLeuAspLysLeuAspSerAl 234
Db 8455 CGAGGACGGCTCAAGTGGGTCAACCATGCTACGAGGGCGCAGGAGCGCAAGCCACCGC 8514
Qy 234 aIleHisLeuCysAsn-----AsnAlaValGlnLysTyrLeuLysAsnAspValGlyAr 252
Db 8515 CGGGGAATGCGCACTGCGCGACTTCGCGCGAAGCACCAAGACGCAAGCCGCGACAT 8574
Qy 252 gSerPro-----LeuLeuProAlaHisAsnMe 261
Db 8575 CACGCCGAGCGCTCAAGTTCCGATCCAGGACGCCATCCGCGTGAAGGCGACGGCAT 8634
Qy 261 t-----TrpThrSerThrArgPheGlnGluTyrLeuGlnArgGlnGlyAlaVa 279
Db 8635 GGACGGCAGCGGACGACCAACATCGACCACTTCATCCAG-----GACGCGGTCAA 8685
Qy 279 lTrpGlySerValIleTyrProSerMetLys----- 289
Db 8686 GTGGGTGTCTTCTGCTACGAGGGGCGAGGCGCGACGCCACCGCGAGGATGCGACGC 8745
Qy 290 ----LysAlaIleAlaHisAlaMetLysValAlaGlnAspHisValGluProArgLysAs 308
Db 8746 CTGGCGCGCTTCGCCAACAGAGAACGAGACATCTCCCGCGGAGGAGCT 8805
```

```
Qy 308 nSerPheGluLeuTyrGlyAla-----AspPheValLeuGlyArgAspPheArgPro-- 325
Db 8806 CAAGTTTCGCATCCAGGACCGCATGCGCAGCAGAGAAAGATGGGACCGAGCTCCGCCAG 8865
Qy 326 -----TrpLeuIleGluLeuAsn----- 331
Db 8866 CAATCGCATCGAAGACCACATCGGGCGGCTACGGCTGGCTCATCTACCTCAACCGAGC 8925
Qy 332 -----SerSerProThrMetHisProSer----- 339
Db 8926 CGGGCCCTCGCAGCCCTCGAGCGCAGCGCCAAACGAGCTGACGAATGGAGAAGTT 8985
Qy 340 -----ThrPro---ValThrAlaGlnLeuCysAlaGlnValGl 351
Db 8986 CGCGCAAAAGAGATTGAAGGAGCGCTGAGATGACGAGGAGGAGCTCCGGTACTACAT 9045
Qy 351 nGluAspThrIleLysValAlaValAspArgSerCysAspIleGlyAsnPheGluLeuLe 371
Db 9046 CCTGGACAGCTCCGACCGCATGCGCAACAGTT--GACGGCGCGCCCTCGGCTCGC 9103
Qy 371 uTrpArgGlnProValValGluProProPheSerGlySerAspLeuCysValAlaGl 391
Db 9104 GTTCAGAGCGAGCGCAGGAAGGCCCAAT-GCCTCAGCTCGCGCTTGTCAAGGCC-- 9160
Qy 391 yValSerValArgAlaArg---ArgGlnValLeuProValCysAsnLeuLysAlase 410
Db 9161 -----TTGAAGCGTTCGCGGACAGCGCTTCGCGCCCTGTCACAGGATGGCTC 9213
Qy 410 rAlaSerLeuLeuAspAlaGlnPro-LeuLysAlaArgGlyProSerAlaMetProAspP 430
Db 9214 C---TGAAGTTCGCGCGCGCGCTCGTGGAGCATCCGACCATCGCGTTACCGCTC 9270
Qy 430 roAla-----G 432
Db 9271 CAGCAACCCAGTCCCAAGGGCGGCTCTCCACTCGGTGCGGTCTTCTCCGTC 9330
Qy 432 lngly-----ProProSerProAlaLeuGlnArgAspLeuGlyLeuLysGluGly 450
Db 9331 AGGGCGCGCTCGCGCAGCTCCGCGCCCATGTCTGAAGAGGGTCTGTCTTCATCAGGTGGGCTC 9390
Qy 450 ly----- 450
Db 9391 GATCGTGTGGACAGCTCCGGGACATCTTCGACCGTCCCGGTGATGAATCCCGCTG 9450
Qy 451 -----LeuProLeuAlaLeuLeuAlap 458
Db 9451 GTGCGACCGCGCAGCCCGCTCTTCTGGAAGAGGGTCTGTCTTCATCAGGTGGGCTC 9510
Qy 458 roLeuArgGlyAlaAlaGluSerGlyAlaAlaGlnProThrArgThrLysAlaAlaG 478
Db 9511 CAACACGTCCGCGCGCGCGGACGCGGATGACCGGAGTACAGCGTCTGTCGCTCCAG 9570
Qy 478 lyLys-----ValGluLeuProAlaC 485
Db 9571 GAGTGGCGCGCAACCGCGGGCCCTTGGCGGCACTCGCGTCAATGCGACTTCTCTGCTT 9630
Qy 485 ysProCysArgHisValAspSerGlnAlaProAsnThrGlyVal----- 499
Db 9631 CGCGCTGC-----AGTTGCTCTCCGATAGCGGTGACGTCTCAGTCCAGGTC 9678
Qy 500 -----ProValAlaGlnProAlaLysSerTrpAspProAsnGlnL 513
Db 9679 CTCGGAGAAGCGCTGGCGACCTGTGCTCCAGGTGGGTGATTGGCTTCCACCCCGA-- 9737
Qy 513 euAsnAlaHisProLeuGluProValLeuArgGlyLeu-----LysThrAlaGluGlyA 531
Db 9738 -----ACGGGCCAGCTCCACTTGGCGGAGTGGACGCTGACCAACCGGTTTCGACGCGC 9792
Qy 531 laLeuArg-----ProProGlyGlyLysGly 540
Db 9793 GGAGACCGCTTCCCATCATCGGTTCATCCGGTTCGCGAGCGCGCTGGA 9839
```

RESULT 12

US-09-248-796A-500
; Sequence 500, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:

; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132

; CURRENT APPLICATION NUMBER: US/09/248,796A
; PRIOR FILING DATE: 1999-02-12

; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13

; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13

; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 500

; LENGTH: 780
; TYPE: DNA

; ORGANISM: Candida albicans
US-09-248-796A-500

Alignment Scores:

Pred. No.:	0.000159	Length:	780
Score:	151.00	Matches:	53
Percent Similarity:	44.0%	Conservative:	50
Best Local Similarity:	22.6%	Mismatches:	81
Query Match:	5.3%	Indels:	50
DB:	3	Gaps:	9

US-10-635-977-2 (1-541) x US-09-248-796A-500 (1-780)

QY	132	AspIleAspGlyLeuArgAsnIleTrpIleLeuLysProAlaAlaLysSerArgGlyArg	151
DB	58	GAATTAATAAGAGAGAAAATTGGATTGTTGAAACCTAGTATGATGATGAAGTCAA	117
QY	152	AspIle-----ValCysMetAspArgValGluGluLeu-----	163
DB	118	GGTATAAGAAATTTTAAACATCGATCAATTGCAAGAAATTTTCAACTCATTTTGAAGAA	177
QY	164	-----GluLeuAlaAlaAspHisPro-----Leu	172
DB	178	AACGATGAGATGAAGAGATGAGGAGGATAGACGAGAGATAACGGTATAATTTTG	237
QY	173	SerArgAspAsnLysTrpValGlnLysTyrIleGluThrProLeuLeuLysCysAsp	192
DB	238	TCTCAACTTCGTCATTATTATTGTTCAAGAATAATAGATTAATCCCTTGTACTA-----	291
QY	193	ThrLysPheAspIleArgGlnTrpPheLeu-----ValThrAspTrpAsnProLeu	209
DB	292	TCAAGATATGACCATAGAAATTCATTTAGCTACTTATGTTGTTGTTGGTGAATTG	351
QY	210	ThrIleTrpPheTyrLysGluSerTyrLeuArgPheSerThrGlnArgPheSerLeu---	228
DB	352	AAAGTATTGTTTATAAGAAATGTTTGTACCTTATTGCTGTGTAACCATATAAATTCCT	411
QY	229	-----AspLysLeuAspSerAlaIleHisLeuCysAsnAsnAlaValGlnLys	244
DB	412	GGTGACGAGGATGAAGTAGTATCTTTACGACGACACTTAAACACACTTGTGTTACAGAA	471
QY	245	TyrLeuLysAsnAspValGlyArgSerProLeuLeuProAlaHisAsnMetTrpThrSer	264
DB	472	-----AATGAGGATCCGTTGGTG-----	489
QY	265	ThrArgPheGlnGluTyrLeuGlnArgGlnGlyAlaValTrpGlySerValIle	284
DB	490	-----GTACCATTTTGAAGAAATTAACAAGTCTTCTGTGATAATGATAAATAATTGTG	540
QY	285	TyrProSerMet-----LysLysAlaIleAlaHisAlaMetLysValAlaGln	300
DB	541	TTTGAACAAATTTGTGATATTACAAAGAAATTTTAAAGCTGTCTACAGTGTGCAAA	600
QY	301	AspHisValGluProArgLysAsnSerPheGluLeuTyrGlyAlaAspPheValLeuGly	320

Db 601 ATGAATTTTCAACCAATAATAATCCATTGAGATATTGGTGTAGATTCTTGGTGAAC 660

QY 321 ArgAspPheArgProTrpLeuIleGluLeuAsnSerSerPro 334

Db 661 AGTGATTTTCTCGTTAAATCTATTAGAAGTTAATTCATACCCT 702

RESULT 13

US-09-252-991A-12848

; Sequence 12848, Application US/09252991A
; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 12848

; LENGTH: 3027

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-12848

Alignment Scores:

Pred. No.:	0.0742	Length:	3027
Score:	132.00	Matches:	149
Percent Similarity:	34.5%	Conservative:	72
Best Local Similarity:	23.2%	Mismatches:	241
Query Match:	4.6%	Indels:	180
DB:	3	Gaps:	34

US-10-635-977-2 (1-541) x US-09-252-991A-12848 (1-3027)

QY 16 SerArgSerArgSerLysProArgAspGlnArgGluAlaGlySerSerAspLeu 35

Db 1258 GCTGATCCGCGCCCTCGCCCGCAGCCGACGAGCAACTGGTGTCTGCG----- 1311

QY 36 SerSerArgGlnAspAlaGluAsnAlaLysLeuArgGly-LeuPro---GlyG1 54

Db 1312 -----CCTGGAACTCAACCCCGCGCTGCGCTGGCGCTGCCCTTCGGCG 1356

QY 54 nLeuValAspIleAlaCysLysValCysGlnAlaTyrLeuGlyGlnLeuGluHisGlu-- 73

Db 1357 CCTGAAGCCGCTGTGGCGCGCTTTCGACTTCTACCTGGCGACCCCGAACCGCAGCG 1416

QY 74 -----AspIleAspThrSerAlaAspAlaValGluAspLeuTh 86

Db 1417 CAAGCTGCGCTGCGCGCACCGGACGCGCGCTGCGCGAT---CTCGACGAGTTGCC 1473

QY 86 xGluAlaGlnTrpGluAspLeuThrGlnGlnTyr-Tyr-SerLeuValHisGlyAspAlaPh 106

Db 1474 GCTGCGC---TGGGAAGGC-----GGCGAC----- 1495

QY 106 eileSerAsnSerArgAsnTyrPheSerGlnCysGlnAlaLeuLeuAsnA-gileThrSe 126

Db 1496 -----AACCTGCGGACTTCGCGCGCTACCGCGCTACCGAGTTTCAGGCTCGT-- 1540

QY 126 rValAsnProGlnThrAspIleAspGlyLeuArgAsnIleTrpIleLeuLysProAlaAl 146

Db 1541 -----CCGCGACCCCGCGCGAAGCGCTGCGC-----GC 1569

QY 146 aLysSerArgGlyArgAspIleValCysMetAspArgValGluGluLeuLeuGluLeu-- 165

Db 1570 CGAACTCGGCGCCCTACCGAGCTCGAGGCGCTGAGCTGATGCGAGACCTTCGCGGACTCGA 1629

QY 166 -----AlaAlaAlaAspHisProLeuSerArgAspAsnLysTrpValValG1 181

Db 166 -----AlaAlaAlaAspHisProLeuSerArgAspAsnLysTrpValValG1 181

QY 286 -ProSerMetLysAlaIleAlaHisAlaMetLysValAlaGlnAspHisValGluPr 305
Db 235 CCGCGGGCGGCGCTGCTGACAGCGCG---CGCTGATCATCGCGAGGTGAATA 291
QY 305 oArgLysAsnSerPheGluLeuTyrGlyAlaAspPheValLeuGlyArg----- 321
Db 292 CTTGAAGAGCGAGTTACCGAGTACGGCAGCGAC---GCCGCGGCGCATATCGGCATCTT 348
QY 322 -----AspPheArgProTrpLeuIleGluIleAsnSerSerPr 334
Db 349 CGCCAATACCAACCGCGGTACCGAAATCTCTCGGAATCCTC-----GCCGG 396
QY 334 oThrMetHisProSerThrProValThrAlaGlnLeuCysAlaGlnValGlnGluAspTh 354
Db 397 CTTCTCCGCGCGTGTGTCGCGGGGTACCTGCACCTGCAGGAAACGCTCAGCGCGGACAT 456
QY 354 rIleLysValAlaValAspArgSerCysAspIleGly-AsnPheGluLeuLeuTyrArg 374
Db 457 TGTCCGCGGCGTGTGACGGCTCCACCGACCTCGGCATCATCGCGCGCGGTGGCGC 516
QY 374 lnProVal-ValGluPro-----ProProPheSerGlySerAspLeuCys 388
Db 517 CACCGCGCTGCAGATCTCATTTACAGACCGCGCGCTGCTGATGGCGT----- 567
QY 389 ValAlaGlyValSerValArg-----AlaArgGlnValLeu 402
Db 568 GCCCGCGGCGCATCTGCTGCGCGGACAGCGGATACCTACGCGACACCTGCAGTT 627
QY 403 ProValCysAsnLeuLysAlaSerAlaSerLeuLeuAspAlaGlnProLeuLysAlaArg 422
Db 628 CCAGCA-----TATCGGCTCGCCGCGGCGC 654
QY 423 GlyProSerAlaMetProAspProAlaGlnGlyProProSer----- 436
Db 655 CACCTCGACGCTCTCTGACGAGCGGTGAGCAGATGGCGAGCAGCTATCGTGGC 714
QY 437 ---ProAlaLeuGln-----ArgAspLeuGlyLeuLysGluGluLys 449
Db 715 CATCGAGTCTCCAGCTTCGAGCGCATCTGCGGATGATCGAGGCGCGCTCGCATCGG 774
QY 450 GlyLeuProLeuAlaLeuAlaProLeuArgGlyAlaAlaGluSerGlyGly----- 467
Db 775 CTTGATCCGCGAGTCCGCGCGCGCCGACAGCGGAGATGAAGCTGGCGACCATCGA 834
QY 468 -----AlaAlaGlnProThrArgThrLysAlaAlaGlyLys 479
Db 835 ACTCGACGAGCCTCGCGGATCCGCGAGCGCAGCTGCTGTGTCGCGCTCGAAGCACT 894
QY 480 ValGluLeuProAlaCysProCysArgHisValAspSerGlnAlaProAsnThrGlyVal 499
Db 895 GCCGAGTTGCTCGC-----GGCGTTGATCGAGGAGTTGCGAGGGTGGG--- 939
QY 500 ProValAlaGlnProAlaLysSerTrpAspProAsnGlnLeuAsnAlaHisProLeuGlu 519
Db 940 -----AGACAGCTGTAGGGCGGATAACGCCCTCGCGTGTGTCGCG--- 981
QY 520 ProValLeuArgGlyLeuLysThrAlaGluGlyAlaLeuArgProProGlyGly 538
Db 982 -----GGGCTCGGAGATCGCGGA-----CAACCCCGCAGCGGT 1017

RESULT 15

US-09-252-991A-12970/c
; Sequence 12970, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 12970
; LENGTH: 1251
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-12970

Alignment Scores:

Pred. No.: 0.032 Length: 1251
Score: 129.50 Matches: 94
Percent Similarity: 33.9% Conservative: 42
Best Local Similarity: 23.4% Mismatches: 141
Query Match: 4.5% Indels: 124
DB: 3 Gaps: 22

US-10-635-977-2 (1-541) x US-09-252-991A-12970 (1-1251)

QY 201 PheLeuValThrAspTrpAsnPro-----LeuThrIleTrpPheTyrLysGluSer 217
Db 1236 TTTCGAATTACCAATGGCGGTTCCGATTTCGCCGAAACGCCATGG-----AGT 1189
QY 218 TyrLeuArgPheSerThrGlnArgPheSerLeuAspLysLeuAspSerAlaIleHisLeu 237
Db 1188 GCCCATGCATTTCCGATCTCGCGA-----CCTGGCGCTGTTCATCCA 1147
QY 238 CysAsnAsnAlaValGlnLysTyrLeuLysAsnAspValGlyArgSerProLeuLeuPro 257
Db 1146 TATCCCGGAGTCGCCAGCCT-----GACCCAGGCGCGCGCTTCCT 1099
QY 258 AlaHisAsnMetTrpThrSerThrArgPheGlnGluTyrLeuGlnArgGlnGlyArgGly 277
Db 1098 -----TTCCCGCGCGCGCGCGCGCGCGCTCAAGC 1066
QY 278 Ala-----ValTrp-GlySerValIleTyr----- 285
Db 1065 GCTCGAAGCGCAGCTTGGCAGCGCGCTGTGTACCGCAGCAGCGCGCGCTCGAACTGAC 1006
QY 286 -ProSerMetLysAlaIleAlaHisAlaMetLysValAlaGlnAspHisValGluPr 305
Db 1005 CCGCGCGCGCGCGCTGCTGTACAGCAGCG---CGGCTGATCATCGCGAGGTGAATA 949
QY 305 oArgLysAsnSerPheGluLeuTyrGlyAlaAspPheValLeuGlyArg----- 321
Db 948 CTTGAAGAGCGAGTTACCGAGTACGGCAGCGAC---GCCGCGGCGCATATCGGCATCTT 892
QY 322 -----AspPheArgProTrpLeuIleGluIleAsnSerSerPr 334
Db 891 CGCCAATACCAACCGCGGTACCGAAATCTCTTCGGGAATCCTC-----GCCGG 844
QY 334 oThrMetHisProSerThrProValThrAlaGlnLeuCysAlaGlnValGlnGluAspTh 354
Db 843 CTTCTCCGCGCGTGTGTCGCGGGTACCGTGCAGCTCGAAGAACCTCAGCGCGGACAT 784
QY 354 rIleLysValAlaValAspArgSerCysAspIleGly-AsnPheGluLeuLeuTyrArg 374
Db 783 TGTCCGCGCGTGTGAGCGCTCCACCGACCTCGGCATCATCGCGCGCGGTGGCGC 724
QY 374 lnProVal-ValGluPro-----ProProPheSerGlySerAspLeuCys 388
Db 723 CACCGCGCTGCAGATCTTCGATTTTCAGCAGCGACCGCTGTGTGATGGCGT----- 673
QY 389 ValAlaGlyValSerValArg-----AlaArgGlnValLeu 402
Db 672 GCCCGAGGCGCATCTCGTGGCGCGGACAGCGCGGATACGCTCAGCGACACCTCGAGTT 613
QY 403 ProValCysAsnLeuLysAlaSerAlaSerLeuLeuAspAlaGlnProLeuLysAlaArg 422
Db 612 CCAGCA-----TATCGGCTCGCGGCGGCGC 586
QY 423 GlyProSerAlaMetProAspProAlaGlnGlyProProSer----- 436

```
Db 585 CACCTGACGGCTTCTGTACGAACGGGTGGAGCAGATGGGCGACAGACTATCGTGGC 526
QY 437 ---ProAlaLeuGln-----ArgAspLeuGlyLeuLysGluGluLys 449
    ||| |||||
Db 525 CATCCAGGTCTCCAGCTTCGAGCGGATCTGCCGGATGATCGAGGGCGGCTCGGCATCG 466
    ||| |||||
QY 450 GlyLeuProLeuAlaLeuAlaProLeuArgGlyAlaAlaGluSerGlyGly----- 467
    ||| |||||
Db 465 CGTGATCCCGCAGTCCCGCGCGCGCCACAGCCGACGATGAAGCTGGCGACCATCGA 406
    ||| |||||
QY 468 -----AlaAlaGlnProThrArgThrLysAlaAlaGlyLys 479
    ||| |||||
Db 405 ACTCGAGAGCCCTGGCGGATCCGCGAGCGCAGCTCTGTTACGGACCTGGAAGCACT 346
    ||| |||||
QY 480 ValGluLeuProAlaCysProCysArgHisValAspSerGlnAlaProAsnThrGlyVal 499
    ||| |||||
Db 345 GCCGAGTTGCTGCG-----GGCGTTGATCGAGGAGTTGCAGAGGTGGG--- 301
    ||| |||||
QY 500 ProValAlaGlnProAlaLysSerTrpAppProAsnGlnLeuAsnAlaHisProLeuGlu 519
    ||| |||||
Db 300 -----AGACAGCCTGTAGGGCGGATAACGCCCTGGCGTTGTCCGCG----- 259
    ||| |||||
QY 520 ProValLeuArgGlyLeuLysThrAlaGluGlyAlaLeuArgProProGlyGly 538
    ||| |||||
Db 258 -----GGGCTCCGGAGATCGGCGGA-----CAACCGCCAGGCGGT 223
    ||| |||||
```

Search completed: April 4, 2006, 09:44:02
Job time : 333 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 4, 2006, 09:35:51 ; Search time 5994 Seconds
(without alignments)
4222.858 Million cell updates/sec

Title: US-10-635-977-2
Perfect score: 2854
Sequence: 1 MASSILKVVVSHQSCRSR.....LRGLKTAEGALRPPGKGKS 541

Scoring table:

BLOSUM62	Xgapop 10.0 , Xgapext 0.5
-Q/-abes/ABSSWEB	Ygapop 10.0 , Ygapext 0.5
-DB=EST -QFMT=fastap	Fgapop 6.0 , Fgapext 7.0
-UNITS=bits	Delop 6.0 , Delext 7.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlh
-Q/-abes/ABSSWEB spool/US10635977/runat 04042006 103532 10020/app query.fasta_1
-DB=EST -QFMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOFC=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000 -HOST=abes07
-USER=US10635977 @CGN 1 1 5315 @runat 04042006 103532 10020 -NCPU=6 -ICPU=3
-NO MAP -NEG SCORER=0 -WAIT -DSPLOCK=100 -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST.*

1:	gb_est1.*
2:	gb_est2.*
3:	gb_est3.*
4:	gb_hcc.*
5:	gb_est4.*
6:	gb_est5.*
7:	gb_est6.*
8:	gb_est7.*
9:	gb_g8s1.*
10:	gb_g8s2.*
11:	gb_g8s2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	DB ID	Description
1	2129.5	74.6	1259	10	AY415398 Homo sapi
2	1963.5	68.8	1257	10	AY415399 Pan trogl
3	1755	2969	4	AK030151	Mus muscu
4	1732.5	60.7	2720	4	AK029745 Mus muscu
5	1639	57.4	1265	10	AY415400 Mus muscu
6	1362	47.7	770	7	CR745100 CR745100
7	1052	36.9	2520	4	AK080321

ALIGNMENTS

RESULT 1	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	PUBMED	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT	
AY415398	Homo sapiens	HCM5544 gene, VIRTUAL TRANSCRIPT, partial sequence, genomic survey sequence.	AY415398	AY415398.1	GI:39771357	GSS.	Homo sapiens (human)	1	(bases 1 to 1259)	Clark,A.G., Gnanowski,S., Nielsen,R., Thomas,P., Kejarawal,A., Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B., Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J., Adams,M.D. and Cargill,M.	Inferred nonneutral evolution from human-chimp-mouse orthologous gene trios	Science 302 (5652), 1960-1963 (2003)	14671302	2	(bases 1 to 1259)	Clark,A.G., Gnanowski,S., Nielsen,R., Thomas,P., Kejarawal,A., Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B., Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J., Adams,M.D. and Cargill,M.	Direct Submission	Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA
AY415398	Homo sapiens	HCM5544 gene, VIRTUAL TRANSCRIPT, partial sequence, genomic survey sequence.	AY415398	AY415398.1	GI:39771357	GSS.	Homo sapiens (human)	1	(bases 1 to 1259)	Clark,A.G., Gnanowski,S., Nielsen,R., Thomas,P., Kejarawal,A., Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B., Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J., Adams,M.D. and Cargill,M.	Inferred nonneutral evolution from human-chimp-mouse orthologous gene trios	Science 302 (5652), 1960-1963 (2003)	14671302	2	(bases 1 to 1259)	Clark,A.G., Gnanowski,S., Nielsen,R., Thomas,P., Kejarawal,A., Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B., Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J., Adams,M.D. and Cargill,M.	Direct Submission	Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA

This sequence was made by sequencing genomic exons and ordering

8	1020	35.7	4184	4	HSW808322
9	998	35.0	775	5	BU937372
10	908	31.8	1059	11	DQ045868
11	891	31.2	1023	3	BM808637
12	811	28.4	1022	3	BM808516
13	811	28.4	1059	11	DQ045869
14	792.5	27.8	777	7	CK597795
15	769	26.9	680	7	CJ018705
16	766.5	26.9	864	6	CD515431
17	764.5	26.8	896	6	CD515431
18	754	26.4	690	8	CD515431
19	751	26.3	813	8	CD515431
20	737.5	25.8	3328	4	BC028169
21	731	25.6	693	3	BM717061
22	701	24.6	679	7	CR986757
23	700	24.5	553	8	DN831611
24	698	24.5	762	6	CD627148
25	694	24.3	776	7	CR998661
26	673	23.6	1371	8	DN726875
27	671	23.5	861	7	CK653583
28	659.5	23.1	787	8	DR722973
29	657	23.0	503	7	CK604006
30	644.5	22.6	616	7	CN349424
31	639.5	22.4	616	5	BM740389
32	638	22.4	651	7	CF999072
33	604	21.2	812	6	CD648578
34	588	20.6	680	6	CB854808
35	587	20.6	583	7	CN349423
36	577	20.2	541	8	DN831612
37	560	19.6	514	7	CR746623
38	520	18.2	831	8	DN532464
39	513	18.0	320	1	AA144756
40	490	17.2	724	2	BG916644
41	485	17.0	285	8	DR008134
42	476	16.7	880	5	BX448465
43	475	16.6	429	8	W31827
44	474	16.6	504	3	BQ327883
45	470	16.5	580	3	BP305902


```

REFERENCE 2 (bases 1 to 1257)
AUTHORS Clark,A.G., Gnanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,
        Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
        Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
        Adams,M.D. and Cargill,M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
        Rockville, MD 20850, USA
COMMENT This sequence was made by sequencing genomic exons and ordering
        them based on alignment.
FEATURES
    source      Location/Qualifiers
                1..1257
                /organism="Pan troglodytes"
                /mol_type="genomic DNA"
                /db_xref="taxon:9598"
                <!.>1257
                /locus_tag="HCM5544"
    gene
    ORIGIN
Alignment Scores:
Pred. No.:      1,11e-160      Length:      1257
Score:          1963.50      Matches:    386
Percent Similarity: 77.6%      Conservative: 2
Best Local Similarity: 77.2%      Mismatches: 25
Query Match:      68.8%      Indels:     86
DB:              10          Gaps:        1
US-10-635-977-2 (1-541) x AY415399 (1-1257)
QY 1 MetAlaSerIleLeuLysTrpValValSerHisGlnSerCysSerArgSerArg 20
DB 18 ATGGCATCCAGATCCTCAAGTGGGTGTGTCAGCCACCAGAGCTGCAGCAGGCGGCAG 77
QY 21 SerLysProArgAspGlnArgGluAlaGlySerSerAspLeuSerSerArgGlnAsp 40
DB 78 AGCAAGCCAGGACGACGAGGAGGAGGCGGGAGCAGCGACCTGAGCAGCAGGCGCAGAT 137
QY 41 AlaGluAsnAlaGluAlaLysLeuArgGlyLeuProGlyGlnLeuValAspIleAlaCys 60
DB 138 GCTGAATATGTGTAGGCAAAAGCTCAGGGGCTCCCGGGGCGAGCTTGTGACGTCGCTGC 197
QY 61 LysValCysGlnAlaTyrluGlyGlnLeuGluHisGluAspIleAspThrSerAlaAsp 80
DB 198 AAGGTGTCCAGGCGCTACTCGGGCAGCTGGAGCATGAGGACATGCACACGTCAGCAGAT 257
QY 81 AlaValGluAspLeuThrGluAlaGluTrpGluAspLeuThrGlnGlnTyrluSerLeu 100
DB 258 GCCGTGGAGGACCTCAGTCAGGCGAGTGGAGGACCTGACCCAGCAGTACTTCCCTC 317
QY 101 ValHisGlyAspAlaPheIleSerAsnSerArgAsnTyrluPheSerGlnCysGlnAlaLeu 120
DB 318 GTTCANNNNNATGCTTTTCATCTCCAATTCAGAAATTAATTTTCGNANNCCAGGCTCTG 377
QY 121 LeuAsnArgIleThrSerValAsnProGlnThrAspIleAspGlyLeuArgAsnIleTrp 140
DB 378 CTGAANNNNNTCANNNTNTNNNCNTNAGACGGACATTCGCGGNNCCGGAACATCTCG 437
QY 141 IleIleLysProAlaAlaLysSerArgGlyValArgAspIleValCysMetAspArgValGlu 160
DB 438 ATTATAAGCCCGCGCCAGTCCCGGGCGCGAGACATAGTGTGATGACACCGGTGGAG 497
QY 161 GluIleLeuGluLeuAlaAlaAspHisProLeuSerArgAspAsnLysTrpValVal 180
DB 498 GAGATCCTGGAGCTGGCAGCTGCAGACCCCTCTTCCAGGGACACAACTGGGTGTC 557
QY 181 GlnLysTyrluGluThrProLeuLeuIleCysAspThrLysPheAspIleArgGlnTrp 200
DB 558 CAGAAGTACATCGAGACCGCTGCTCATCTATGACACCAAGTTCGACATCAGACAGTGG 617
QY 201 PheLeuValThrAspTrpAsnProLeuThrIleTrpPheTyrluGluSerTyrluArg 220
DB 618 TTCCTCGTCAGGAGTGAACCCCTGACCATCTGGTCTTCAAGGAGAGGTACTTTCGG 677

```

```

QY 221 PheSerThrGlnArgPheSerLeuAspLysLeuAspSerAlaIleHisLeuCysAsnAsn 240
DB 678 TTCTCAACACACGCGCTTCTCCCTGGACAAGCTGGACGCGCCATCCACCTGTGTCAACAC 737
QY 241 AlaValGlnLysTyrluLysAsnAspValGlyArgSerProLeuLeuProAlaHisAsn 260
DB 738 GCCGTCCAGAAGACCTGAAGAATATGTGGGCCGACGCCCTCTGCTGCCCCACACACAC 797
QY 261 MetTrpThrSerThrArgPheGlnGluTyrluLeuGlnArgGlnGlnArgGlyAlaValTrp 280
DB 798 ATGTGACACGCGCCAGGTTCCAGAGTACCTGTCAGCGCCAGGCGCGCGCTGTGG 857
QY 281 GlySerValIleTyrluProSerMetLysLysAlaIleAlaHisAlaMetLysValAlaGln 300
DB 858 GGCNNNGTCATCTACCCGTCCATGAAGAGGCGCATCGCCACACCATGAAGGTGGCCAG 917
QY 301 AspHisValGluProArgLysAsnSerPheGluLeuTyrluGlyAlaAspPheValLeuGly 320
DB 918 GACCACGTGGAGCCTCGCAAGACAGCTTACGCTCTACGGGGCTGACTTCGTCTCTGGG 977
QY 321 ArgAspPheArgProTrpIleGluIleAsnSerSerProThrMetHisProSerThr 340
DB 978 AGGACTTCAGNCCCTGGCTGATCGAGATCAATTCACGCCCCACCATGCACCCGTCACG 1037
QY 341 ProValThrAlaGlnLeuCysAlaGlnValGlnGluAspThrIleLysValAlaValAsp 360
DB 1038 CCGGTACGCGCCAGCTGTGTGCACAGGTGCAGGAGGACACCATCAAGGTGCGCTGGAC 1097
QY 361 ArgSerCysAspIleGlyAsnPheGluLeuLeuTrpArgGlnProValValGluProPro 380
DB 1098 CGCAGCTGTGACATCGGCAACTTCGAGNCCCTGTGGAGCCTNNNNNGTGTGTGAGCGGCC 1157
QY 381 ProPheSerGlySerAspLeuCysValAlaGlyValSerValArgArgAlaArgGln 400
DB 1158 CCATTACGCGGTGCGACCTCTGTGTGGCGGGCTCAGTAG----- 1198
QY 401 ValLeuProValCysAsnLeuLysAlaSerAlaSerLeuLeuAspAlaGlnProLeuLys 420
DB 1198 ----- 1198
QY 421 AlaArgGlyProSerAlaMetProAspProAlaGlnGlyProProSerProAlaLeuGln 440
DB 1198 ----- 1198
QY 441 ArgAspLeuGlyLeuLysGluGluLysGlyLeuProLeuAlaLeuAlaProLeuArg 460
DB 1198 ----- 1198
QY 461 GlyAlaAlaGluSerGlyGlyAlaAlaGlnProThrArgThrLysAlaAlaGlyLysVal 480
DB 1198 ----- 1198
QY 481 GluLeuProAlaCysProCysArgHisValAspSerGlnAlaProAsnThrGlyValPro 500
DB 1199 ---CTCCGCGCTCCCTCTGTGCGCAGTGGACAGTCAAGGCCCAACACCGGTGTGCTCCC 1255
RESULT 3
LOCUS AK030151 2969 bp mRNA linear HTC 03-APR-2004
DEFINITION Mus musculus adult male testis cDNA, RIKEN full-length enriched
        library, clone:4933401B17 product:hypothetical Tubulin-tyrosine
        ligase containing protein, full insert sequence.
ACCESSION AK030151
VERSION AK030151.1 GI:26326134
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
        Mus musculus
        ORGANISM
        Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
        Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
        Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning

```

JOURNAL PUBMED REFERENCE AUTHORS	Meth. Enzymol. 303, 19-44 (1999) 10349636 2
TITLE	Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome Res. 10 (10), 1617-1630 (2000) 11042159
JOURNAL PUBMED REFERENCE AUTHORS	3 Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsuana,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer Genome Res. 10 (11), 1757-1771 (2000) 11076861
TITLE	4 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium. Functional annotation of a full-length mouse cDNA collection Nature 409, 685-690 (2001)
JOURNAL PUBMED REFERENCE AUTHORS	5 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs Nature 420, 563-573 (2002) 6 (bases 1 to 2969)
JOURNAL PUBMED REFERENCE AUTHORS	Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Hori,P., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T., Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,Y., Tanaka,T., Tomaru,A., Toyota,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y. Direct Submission Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
TITLE	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Please visit our web site for further details. URL:http://genome.gsc.riken.jp/ URL:http://fantom.gsc.riken.jp/ Location/Qualifiers 1. .2969 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="FANTOM,DB:4933401B17" /db_xref="taxon:10090" /clone="4933401B17" /sex="male" /tissue_type="testis" /clone_lib="RIKEN full-length enriched mouse cDNA library" /dev_stage="adult"
FEATURES source	

CDS	490..2835 /note="unnamed protein product; hypothetical Tubulin-tyrosine ligase containing protein (InterPro IPR004344, evidence: InterPro) putative" /codon_start=1 /protein_id="BAC26811.1" /db_xref="GI:26326135" /translation="MENGERRKSLSTLSDGDHKEKNLKGIPQDLSSSPKLDRYKIA ROLTEKAIKERKIFSIYGHYPVIRATLRKMGVEKKNFPKALQNLGDSKAEYKE NOEIALERFDDIHDVMSRLVKNIEFYLLWTKRDVVDVHSLTCDQMLNHYGKTASTFT KIGLCNKRSLSPWYVOANPNTFFPCYGLCTESEKQELDDFRRTVAASLKLWVLUHQ NYSVKVGRSKKEEAKNDSPKDPENPLKLSLQGVVDVDTACTVCQAYLGLQLEHE DIDVSEASTEALSEBENDLTOQYLLVHGNASITDSKSYFAQCOALLKSISSNPQT EDJGRNIWIKPAKSGRGRDIVCDVRNLSLVAADSOTTCKDNKVVQKYIETPML IYDTFEDIQWFLVTDNPLTTFWYKESYLRESTORFSLDKLDSAHLCNNSTORLKL NDKRSPLGPDCHNMMTSTRFYLOKRGGRGTWGSIIYPSMKRAVTAQVADHVEA RNSPFLYGADFILGRDFPWLIEINSSPTFTVTAOLCAQOVEDITKVVDVRLK DRNCIDGNFELTWQPAVELPFGNDLCEGISVKAKKOMPPTASVGRSKPTIYAEV PKVRSARALMETVIRPRTTVRQDWKREAEVLSTTWSMPMDAEVGRGAKPTIYAEV NDYOHVDNKSXKSGYTRVQSSKVPGLVLTSAOHPALFAOTMKTPTQMTSSPPPTASGNH RDSFPCPIVFEELMLHPNSQRRPSLQSRAGQWIRGIP"
ORIGIN	US-10-635-977-2 (1-541) x AK030151 (1-2969) Alignment Scores: Pred. No.: 6,53e-142 Length: 2969 Scores: 1755.00 Matches: 348 Percent Similarity: 76.6% Conservative: 72 Best Local Similarity: 63.5% Mismatches: 108 Query Match: 61.5% Indels: 21 DB: 4 Gaps: 8 QY 1 MetAlaSerSerIleLeuLysTrpValValSerHisGlnSer---CysSerArgSerSer 19 Db 1105 GTGGCTGCCAGCATCTCTCAAGTGGGTAGTCTCTTACCAGAAATTACTGC-----AGC 1155 QY 20 ArgSerLysProArgAspGlnArgGluGluAlaGlySerSerAspLeuSerSerArgGln 39 Db 1156 AAGTCAAGGGCAAGTAGTAAGAGGAGGAGGCCAAGAACAGTAGTCCGCCAGCCCAAGAAA 1215 QY 40 AspAlaGluAsnAlaGluAlaLysLeuArgGlyLeuProGlyGlnLeuValAspIleAla 59 Db 1216 GATCCTGAAATCTCGACCTCAAGCTACCAAGCTCTCGGAGACAGGTGTGACACACAGCT 1275 QY 60 CysLysValCysGlnAlaTyrlLeuGlyGlnLeuGluHisGluAspIleAspThrSerAla 79 Db 1276 TGCAAGGTGTGCCAGGCTTACCTGGGACAGCTAGAGCATGAAGACATAGACGTATCGGAG 1335 QY 80 AspAlaValGluAspLeuThrGluAlaGluTrpGluAspLeuThrGlnGlnTyrlSer 99 Db 1336 GCCAGCAGCAGGCGCTTGTCTGAGGAGGAATGGAATGACTGACACAGCAGTACTATCTA 1395 QY 100 LeuValHisGlyAspAlaPheIleSerAsnSerArgAsnTyrlPheSerGlnCysGlnAla 119 Db 1396 CTGGTTCATGGCAATGCTTCCATCATTGATTCGAAAGTTACTTTGGCAATGCGCAAGCT 1455 QY 120 LeuLeuAsnArgIleThrSerValAsnProGlnThrAspIleAspGlyLeuArgAsnIle 139 Db 1456 CTGCTGAGTAAGATCAGTTCTGTGAACCTCTCAGACAGAGATCGATGGATCGCGAATC 1515 QY 140 TrpIleIleLysProAlaAlaLysSerArgGlyArgAspIleValCysMetAspArgVal 159 Db 1516 TGGATCATAAAGCTTCAGCAAGTCCCGGGGTGAGATATTGTGTATGGACCGGTGTG 1575 QY 160 GluGluIleLeuGluLeuAlaAlaAspHisProLeuSerArgAspAsnLysTrpVal 179 Db 1576 GAGAACATCTCTGAGTCTGGTGGCGCAGACAGCCAGACCAAGGACCAACAAATGGGTG 1635 QY 180 ValGlnLysTyrlLeuGluThrProLeuLeuIleCysAspThrLysPheAspIleArgGln 199 Db 1636 GTCCAGAAAGTACATCGAGAGCCCAATGCTCATCTACGACACCAAGTTTGACATCAGACAG 1695

Qy	200	TrpPheLeuValThrAspTrpAsnProLeuThrIleLtpPheTyrLysGluSerTyrLeu	219
Db	1696	TGGTTCCTCGTCAACAGCTGGAAATCCCTTAACCATCTGGTTCATAAAGAAAGCTACCTC	1755
Qy	220	ArgPheSerThrGlnArgPheSerLeuAspLysLeuAspSerAlaIleHisLeuCysAsn	239
Db	1756	CGCTTCTCCACACAGCGCTTCTCCCTGGACAAGCTGGACAGCGCATCCACCTGTGTAAAC	1815
Qy	240	AsnAlaValGlnLysTyrLeuLysAsnAspValGlyArgSerProLeuLeuProAlaHis	259
Db	1816	AACTCCATCCAGAGCGCGCTCAAAAATGATAAAGAGCGCATCGCGCTGCTACCTTGTTCAT	1875
Qy	260	AsnMetTrpThrSerThrArgPheGlnGlnLysLeuGlnArgGlnGlyArgGlyAlaVal	279
Db	1876	AACATGTGGACAGCATCTGTTTTCAGAGATPACTCTGCAGAAGAGAGGGCGGAGGAGAACG	1935
Qy	280	TrpGlySerValIleTyrProSerMetLysLysAlaIleAlaHisAlaMetLysValAla	299
Db	1936	TGGGGTAGCATCATCTACCGCTCTATGAAGAAGAGCTGTCAACCAATGCCATGAGGGTAGCC	1995
Qy	300	GlnAspHisValGluProArgLysAsnSerPheGluLeuTyrGlyAlaAspPheValLeu	319
Db	1996	CAGACACACGTAGAAGCGCGCTAAGAACAGCTTCAGAGCTCTACGAGCGCGACTTCATCCTG	2055
Qy	320	GlyArgAspPheArgProTyrLeuIleGluLysAsnSerSerProThrMetHisProSer	339
Db	2056	GGGCGAGACTTCAGCCATGGCTTATCGAGATCAACTCCAGCCGCCACATGACGCCCTCC	2115
Qy	340	ThrProValThrAlaGlnLeuCysAlaGlnValGlnGluAspThrLysValAlaVal	359
Db	2116	ACTCCCGTCAAGCTCAGCTCTGTGCCCAGGTGCAGGAGGACACCATCAAGGTGGTGGTG	2175
Qy	360	-----AspArgSerCysAspIleGlyAsnPheGluLeuLeuTyrArgGlnPro	375
Db	2176	GATCGCAAACTGGAGCCGAAACTGTGACATCGGCAACTTTCAGCTTCTGTGGCGCAGCCT	2235
Qy	376	ValValGluProProPheSerGlySerAspLeuCysValAlaGlyValSerValArg	395
Db	2236	GCGTGGAGTGCACCGCTTCAACGGGTCTGACCTCTGTGTGGAAAGTATCAGCGGTGAAG	2295
Qy	396	ArgAlaArgGlnValLeuProValCysAsnLeuLysAlaSerAlaSerLeuLeuAsp	415
Db	2296	AAAGCCAAAGAAGCAATCGCCCATTCGTCAGCGCTCGGCCCTCTCAGAAATCCTCTTAGAT	2355
Qy	416	AlaGlnProLeuLysAlaArgGlyProSerAlaMetProAspProAlaGlnGlyProPro	435
Db	2356	GCTTCCACCC---AAAGTGGCGAGTGGCCGGCGCCCTGTATGGAAACAGTGTATCAGGCCCTCCA	2412
Qy	436	SerProAlaLeuGlnArgAspLeuGlyLeuLysGluGluLysGlyLeuProLeuAlaLeu	455
Db	2413	AGAACGACCTGGCGAGAGACTGGAAACGGGAAGAAGCAAGGTACTTTCAACCACTTGG	2472
Qy	456	LeuAlaProLeuArgGlyAlaAlaGluSerGlyGlyAlaGlnProThrArgThrLys	475
Db	2473	TCCATGCCAGTAATGGAT---GCAGAGGTTCAGAGGTTCGAGCAAGCCCATCTATGTCATTC	2529
Qy	476	AlaAlaGlyLysValGlnLeuProAlaCysProCysArgHisValAspSerGlnAlaPro	495
Db	2530	GAGGTCAATGACTACCAA-----CATGTAGACAATAAGTCCCCAC	2568
Qy	496	AsnThrGly-ValProValAlaGlnProAlaLysSerTrpAsp---ProAsnGlnLeuAs	514
Db	2569	AAGAGCGGTATACAGAGTCCAAATCAAGTAAAGTCCCAGGGGTAACTTAAACAGTGA	2628
Qy	514	nalAlaHisProLeuGluProValLeuArgGlyLeuThrAlaGluGlyAlaLeuArgPr	534
Db	2629	CAG-CACCCCT---GCCCTGTTTCACACAGACTATGAACCCACACAGATCACCTCGTCTCC	2684
Qy	534	oProProGlyGlyLysGlySer	541
Db	2685	ACCACCCCATCGCTCAGGTAAC	2706

[illegible]

Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site for further details.

URL: <http://genome.gsc.riken.jp/>

URL: <http://fantom.gsc.riken.jp/>

FEATURES

source

```
Location/Qualifiers
1..2720
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:4930524K07"
/db_xref="taxon:10090"
/clone="4930524K07"
/sex="male"
/tissue_type="testis"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
934..2430
/notes="unnamed protein product; hypothetical
Tubulin-tyrosine ligase containing protein
(InterPro IPR004344, evidence: InterPro)
putative"
/codon_start=1
/protein_id="BAC26595.1"
/db_xref="GI:2632582"
/translation="MLNHYGKTSFTTKIGLCLNMRSLPWYVOANPTFFPCYGLCT
ESEKQFLDPRFTVAASILKWLVLHONYCSVKGSKEEAKNSDPSPKKDPENPDL
KLPSLGGVVDYACKVQALYGOLEHEDIDVSEASTEALSEEWNLDLTOQYLLVHGN
ASTLSDSKSPFAOCCALLSKSSVNPOTEDIGRNWIIPAAKSRGRDIVCMRDVENI
LSLVAADSTQTDKNKVVQKYLETPLMIYDTKFDIRQWFLVTDNPLRTIWFYKESYLR
FSTQRLSDLSAHLNNS IORLKNKERSPLLPCHNMTSTSFQYLQKRGSG
TWGSIITPMSKRAVTNMRVAQDHVEARKNSFELYGADFILGRDFPKPLIEINSSPTM
HPSTPTAOLCAQVQDITI KVVVDRLKDRNCIDGNFELLWRQPAVSLPFPNGSDLCVE
GISVKKAKOMPPIASVGLSESLDAPPKVRARALMETVIRPPRTTVRQDWKREAK
GFRVISKRRR"
```

CDS

ORIGIN

Alignment Scores:

```
Pred. No.: 5,3e-140 Length: 2720
Score: 1732.50 Matches: 337
Percent Similarity: 80.7% Conservative: 60
Best Local Similarity: 68.5% Mismatches: 81
Query Match: 60.7% Indels: 15
DB: 4 Gaps: 5

US-10-635-977-2 (1-541) x AK029745 (1-2720)

Qy 1 MetAlaSerSerIleLeuLysTyrValValSerHisGlnSer---CysSerArgSerSer 19
Db 1108 GTGGCTGCCAGCATCTCTCAAGTGGGTAGTCTTCCACCAAGAAATTAAGTCTGC 1158
Qy 20 ArgSerLysProArgAspGlnArgGluGluAlaGlySerSerAspLeuSerSerArgGln 39
Db 1159 AAGTCAAGGGCAGAGATAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1218
Qy 40 AspAlaGluAsnAlaGluAlaLysLeuArgGlyLeuProGlnGlnLeuValAspIleAla 59
Db 1219 GATCCTGAAATCTCTGACCTCAAGCTACCAAGCTCTCTCGGACACAGGTTGTAGACACAGCT 1278
Qy 60 CysLysValCysGlnAlaTyrLeuGlyGlnLeuGluHisGluAspIleAspThrSerAla 79
Db 1279 TGCAAGGTGTGCCAGGCTACCTCGGACAGCTAGACATGAAGACATAGACGTATCGGAG 1338
Qy 80 AspAlaValGluAspLeuThrGluAlaGluTyrGluAspLeuThrGlnTyrTyrSer 99
Db 1339 GCCAGCACCGAGGCTTGTCTGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1398
Qy 100 LeuValHisGlyAspAlaPheIleSerAsnSerArgAsnTyrPheSerGlnCysGlnAla 119
Db 1399 CTGTTTCATGGCAATGCTTCCATCACTGATTCGAAAAAGTTACTTTGGCGCAATGCCAAGCT 1458
Qy 120 LeuLeuAsnArgIleThrSerValAsnProGlnThrAspIleAspGlyLeuArgAsnIle 139
Db 120 LeuLeuAsnArgIleThrSerValAsnProGlnThrAspIleAspGlyLeuArgAsnIle 139
```

```
Db 1459 CTGCTGAGTAGAATCAGTTCTGTGAACCTTCAGACAGAGATCGATGGGATCCGGAACATC 1518
Qy 140 TrpIleIleLysProAlaLysSerArgGlyArgAspIleValCysMetAspArgVal 159
Db 1519 TGGATCATAAAGCCTGCAGCCAAAGTCCCGGGTTCGAGATATTGTGTATGACCGGTGTG 1578
Qy 160 GluGluIleLeuGluLeuAlaAlaAspHisProLeuSerArgAspAsnLysTyrVal 179
Db 1579 GAGAACATCTGAGTCTGGTGGCCGACAGACAGCCAGACCAAGGACCAACAATGGGTG 1638
Qy 180 ValGlnLysTyrIleGluThrProLeuLeuIleCysAspThrLysPheAspIleArgGln 199
Db 1639 GTCCAGAAGTACATCGAGACCCCAATGCTCATCTACGACACCAAGTTTGCATCAGACAG 1698
Qy 200 TrpPheLeuValThrAspTyrAsnProLeuThrIleTyrPheTyrLysGluSerTyrLeu 219
Db 1699 TGGTTCCTCGTCACAGACTGGAATCCCTTAACCAATCTGTTCTTACAAAGAAAGCTACCTC 1758
Qy 220 ArgPheSerThrGlnArgPheSerLeuAspLysLeuAspSerAlaIleHisLeuCysAsn 239
Db 1759 CGCTTCTCCACAGCGCTTCTCCCTGGACAGCTGGACAGCGCCATCCACCTGTGTAAAC 1818
Qy 240 AsnAlaValGlnLysTyrLeuLysAsnAspValGlyArgSerProLeuLeuProAlaHis 259
Db 1819 AACTCCATCCAGAGCGCCCTCAAAATATGATAAAGAGCGCAGTCCGCTGTCTTCTCAT 1878
Qy 260 AsnMetTyrThrSerThrArgPheGlnGluTyrLeuGlnArgGlnGlyValArgGlyAlaVal 279
Db 1879 AACATGTGACACGACACTCGTTTCCAGGAGTAGTCTCGAAGAGAGGGGCCGAGAGAACG 1938
Qy 280 TrpGlySerValIleTyrProSerMetLysLysAlaIleAlaHisAlaMetLysValAla 299
Db 1939 TGGGGTAGCATCATCTACCGCTATGAAAGAGCTGTCCACCAATGCCATGAGGGTAGCC 1998
Qy 300 GlnAspHisValGluProArgLysAsnSerPheGluLeuTyrGlyAlaAspPheValLeu 319
Db 1999 CAGGACCACTAGAGCCGCTAAGAACAGCTTTCAGCTTCTACGAGCGCAGCTTTCATCCTG 2058
Qy 320 GlyArgAspPheArgProTyrLeuIleGluIleAsnSerSerProThrMetHisProSer 339
Db 2059 GGGCGAGACTTCAAGCCATGGCTTATCGAGATCAACTCCAGCCGCCACCTGACCCCTCC 2118
Qy 340 ThrProValThrAlaGlnLeuCysAlaGlnValGlnGluAspThrIleLysValAlaVal 359
Db 2119 ACTCCCGTCACAGCTCAGCTCTGTGCCAGGTCCAGAGGACACCATCAAGGTGGTGTG 2178
Qy 360 -----AspArgSerCysAspIleGlyAsnPheGluLeuLeuTyrArgGlnPro 375
Db 2179 GATCGCAAACTGGACCGMAACTGTGACATCGGCAACTTTGAGCTTCTGTGGCGCAGCCT 2238
Qy 376 ValValGluProProPheSerGlySerAspLeuCysValAlaGlyValSerValArg 395
Db 2239 GCCGTGGAGCTGCCCGTTCAACGGGTCTGACCTCTGTGTGAAGGTATCAGCGTGAAG 2298
Qy 396 ArgAlaArgArgGlnValLeuProValCysAsnLeuLysAlaSerAlaSerLeuLeuAsp 415
Db 2299 AAAGCCAGAGCAAGATGCCCGCCATTCGACGGTCCGCGCTCTCAGAAATCACTCTTAGAT 2358
Qy 416 AlaGlnProLeuLysAlaArgGlyProSerAlaMetProAspProAlaGlnGlyProPro 435
Db 2359 GCTCCACCC---AAAGTGGGAGTGGCCGGCGCTGATGGAAACAGTGTATCAGGCTTCCA 2415
Qy 436 SerProAlaLeuGlnArgAspLeuGlyLeuLysGluGluLysGlyLeuProLeuAlaLeu 455
Db 2416 AGAACGACCGTGGCGGAGACTGGAAACGGGAAAGAGC-AAAGGGCTTTCGTGTGATCTC 2474
Qy 456 LeuAlaProLeuArgGlyAlaAla-----GluSerGlyGlyAlaAlaGln 470
Db 2475 CAGCGCGGTGCTGAGTTCTGTGGCTTCGCGCCGCCCGCTTGGCGCTTGGCGAGCTC 2534
Qy 471 ProThrArgThrLysAlaAlaGlyLysValGluLeu 482
Db 2535 TCCCTGCGCTTGTATGTCGCGAGGCGCTTCTGTCTG 2570
```

```

RESULT 5
AY415400      1265 bp      DNA      linear      GSS 12-DEC-2003
LOCUS      Mus musculus HCM5544 gene, VIRTUAL TRANSCRIPT, partial sequence,
DEFINITION      genomic survey sequence.
ACCESSION      AY415400
VERSION      AY415400.1 GI:39771359
KEYWORDS      GSS.
SOURCE      Mus musculus (house mouse)
ORGANISM      Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1265)
Clark, A.G., Gnanowski, S., Nielson, R., Thomas, P., Kejariwal, A.,
Todd, M.A., Tanenbaum, D.M., Civeello, D.R., Lu, F., Murphy, B.,
Ferriera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,
Adams, M.D. and Cargill, M.
Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
Science 302 (5652), 1960-1963 (2003)
14671302
2 (bases 1 to 1265)
Clark, A.G., Gnanowski, S., Nielson, R., Thomas, P., Kejariwal, A.,
Todd, M.A., Tanenbaum, D.M., Civeello, D.R., Lu, F., Murphy, B.,
Ferriera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,
Adams, M.D. and Cargill, M.
Direct Submission
Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
This sequence was made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
source      1..1265
            /organism="Mus musculus"
            /mol_type="genomic DNA"
            /db_xref="taxon:10090"
            <1..>1265
            /locus_tag="HCM5544"
gene
ORIGIN
Alignment Scores:
Pred. No.:      2,53e-132      Length:      1265
Score:      1639.00      Matches:      305
Percent Similarity:      87.4%      Conservative:      43
Best Local Similarity:      76.6%      Mismatches:      42
Query Match:      57.4%      Indels:      8
DB:      10      Gaps:      3
US-10-635-977-2 (1-541) x AY415400 (1-1265)
QY      1      MetAlaSerIleLeuYstPValValSerHisGlnSer---CysSerArgSerSer 19
Db      18      GTGGCTGCAGCATCTCAAGTGGGTAGTCTTCCACGAAATTACTGC-----AGC 68
QY      20      ArgSerLysProArgAspGlnArgGluGluAlaGlySerSerAspLeuSerSerArgGln 39
Db      69      AAAGTCAAGGGCAAGAGTAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 128
QY      40      AspAlaGluAunAlaGluAlaLysLeuArgGlyLeuProGlyGlnLeuValAspIleAla 59
Db      129      GATCTGAAATTCCTTGACCTCAAGCTTACCAAGCTCTCGGAGACAGGTTGTAGACACAGT 188
QY      60      CysLysValCysGlnAlaTyrluGlyGlnLeuGluHisGluAspIleAspThrSerAla 79
Db      189      TGCAGAGGTGTCAGGCGCTTACCTGGGACAGCTAGAGCATGAGACATAGACGTATCGGAG 248
QY      80      AspAlaValGluAspLeuThrGluAlaGluTrpGluAspLeuThrGlnGlnTyrlSer 99
Db      249      GCCAGACCGAGCGCTTCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 308
QY      100      LeuValHisGlyAspAlaPheIleSerAsnSerArgAsnTyrlPheSerGlnCysGlnAla 119

```

```

Db      309      CTGGTTTCATGGCAATGCTTCCATCATGATTGAAAAAGTTACTTTGCCGAATGCCAAGCT 368
QY      120      LeuLeuAsnArgIleThrSerValAsnProGlnThrAspIleAspGlyLeuArgAsnIle 139
Db      369      CTGCTGAGTAAGATCAGTTCTGTGAACCTCAGACAGAGATCAGTGGATCGGAACATC 428
QY      140      TrpIleIleLysProAlaLysSerArgGlyArgAspIleValCysMetAspArgVal 159
Db      429      TGGATCATAAAGCCTGCAGCAAGTCCCGGGTTCAGAGATATTGTGTGTATGACCGTGTG 488
QY      160      GluGluIleLeuGluLeuAlaAlaAspHisProLeuSerArgAsnIleYstPVal 179
Db      489      GAGAACATCTCTGAGTCTGTGTGGCCGACAGACCCAGACCAACAAAGACCAAAATGGGTG 548
QY      180      ValGlnLysTyrlleGluThrProLeuLeuIleCysAspThrLysPheAspIleArgGln 199
Db      549      GTCCAGAGTATACGAGACCCCAATGTCTCATCAGACACCCAGTTTGACATCAGACAG 608
QY      200      TrpPheLeuValThrAspTrpAsnProLeuThrIleTrpPheTyrlsGluSerTyrlLeu 219
Db      609      TGGTTCTCTCGTCACAGACTGGAATCCCTTAACCATCTGTGTTCTACAAAGAAAGTACCTC 668
QY      220      ArgPheSerThrGlnArgPheSerLeuAspLysLeuAspSerAlaIleHisLeuCysAsn 239
Db      669      CGCTTCTCCACACAGCGCTTCTCCCTGGACAAAGCTGGACAGCGCCATCCACCTGTGTAAC 728
QY      240      AsnAlaValGlnLysTyrlleLysAsnAspValGlyArgSerProLeuLeuProAlaHis 259
Db      729      AACTCCATCCAGAGCGGCTCAAAATGATAAAGAGCGAGTCCGCTGTCTGTCTGTCTAT 788
QY      260      AsnMetTrpThrSerArgPheGlnGluTrpLeuGlnArgGlnGlnArgGlyAlaVal 279
Db      789      AACATGTGGAGCCAGCACTCGTTTCCAGGAGTACCTTCAGAGAGGAGGAGGAGGAGGAG 848
QY      280      TrpGlySerValIleTyrlProSerMetLysLysAlaIleAlaHisAlaMetLysValAla 299
Db      849      TGGGGTAGCATCATCTACCCGCTCTATGAAAAGAGCTGTACCAATGCCATGAGGGTAGCC 908
QY      300      GlnAspHisValGluProArgLysAsnSerPheGluLeuTyrlGlyAlaAspPheValLeu 319
Db      909      CAGGACCACTAGAGCCCGGTAAAGAACAGTCTCGAGCTCTACGAGGCCACATTCATCCTG 968
QY      320      GlyArgAspPheArgProTrpLeuIleGluIleAsnSerSerProThrMetHisProSer 339
Db      969      GGGGAGAGCTTCAAGCCATGGCTTATCGAGATCACTCCAGCCCCCAGCCATGACCCCTCC 1028
QY      340      ThrProValThrAlaGlnLeuCysAlaGlnValGlnGluAspThrIleLysValAlaVal 359
Db      1029      ACTCCCGTCACAGCTCAGCTCTGTGCCAGGTGCGAGGAGGACCAATCAAGGTGTGTGTG 1088
QY      360      -----AspArgSerCysAspIleGlyAsnPheGluLeuTrpArgGlnPro 375
Db      1089      GATCGAAACTGGACCGGAAACTGTGACATCGCAACTTGTAGCTTCTGTGGCGGAGCCT 1148
QY      376      ValValGluProProPheSerGlySerAspLeuCysValAlaGlyValSer 393
Db      1149      GCCGTGGAGCTGCCACCGTTCACCGGTCTGACCTCTGTGTGGAAGGTATCAGC 1202
RESULT 6
CR745100      770 bp      mRNA      linear      EST 21-DEC-2004
LOCUS      CR745100 Soares testis NHT Homo sapiens cDNA clone IMAGE998M164174
DEFINITION      ; IMAGE:1645239"5", mRNA sequence.
ACCESSION      CR745100
VERSION      CR745100.1 GI:51667573
KEYWORDS      EST.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
REFERENCE      1 (bases 1 to 770)

```

AUTHORS

Ebert, L., Heil, O., Hennig, S., Korn, B., Neubert, P., Partsch, E.,

TITLE

I.M.A.G.E. cDNA Clone Collection

JOURNAL

Unpublished (2004)

COMMENT

Contact: Inge Arlart

RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH

Heubnerweg 6, D-14059 Berlin, Germany

E-mail: www.rzpd.de

RZPD; IMAG998M164174.

RZPDLIB; I.M.A.G.E. cDNA Clone Collection;

Contact: Inge Arlart

RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH

Heubnerweg 6, D-14059 Berlin, Germany

Tel: +49 30 32639 100

Fax: +49 30 32639 111

www.rzpd.de

This clone is available royalty-free from RZPD;

contact RZPD (clone@rzpd.de) for further information. Seq primer:

M13r, Primer sequence: TTTCACAGGAACAGCTATGAC.

FEATURES

Location/Qualifiers

source

1..770

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAG998M164174 ; IMAGE:1645239"

/sex="male"

/lab_host="DH10B"

/clone_lib="Soares testis NHT"

/notes="Vector: pT73D-Pac (Pharmacia) with a modified

polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA

was prepared from mRNA obtained from Clontech

Laboratories, Inc., and primed with a Not I - oligo(dT)

primer [5].

TGTTACCAATCTGAGTGGAGCGCGCCCAATTTTTTTTTTTT 3'.

Double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not I

and Eco RI sites of the modified pT73 vector. Library

went through one round of normalization to Cot5, and was

constructed by Bento Soares and M. Fatima Bonaldo. "

ORIGIN

Alignment Scores:

Pred. No.: 1..94e-108 Length: 770

Score: 1362.00 Matches: 254

Percent Similarity: 99.2% Conservative: 0

Best Local Similarity: 99.2% Mismatches: 2

Query Match: 47.7% Indels: 0

DB: 7 Gaps: 0

US-10-635-977-2 (1-541) x CR745100 (1-770)

Qy	85	LeuThrGluAlaGluTrpGluAspLeuThrGlnGlnTrpTyrSerLeuValHisGlyAsp	104
Db	3	CTCACTGAGCCGAGTGGAGGACCTGACCCAGCAGTACTACCTCGTTCTATGGCGAT	62
Qy	105	AlaPheIleSerAsnSerArgAsnTyrPheSerGlnCysGlnAlaLeuLeuAsnArgIle	124
Db	63	GCATTTCATCTCCAATTCAAGAAATTAATTTTCGCAAGTGCAGGCTCTGCTGAATAGATC	122
Qy	125	ThrSerValAsnProGlnThrAspIleAspGlyLeuArgAsnIleTrpIleIleLysPro	144
Db	123	ACGCTCTGTGAACCTTCAGAGGACATGACGGGCTCCGGAACTCTGGATTATTAAGCCC	182
Qy	145	AlaAlaIysSerArgGlyArgAspIleValCysMetAspArgValGluGluLeuGlu	164
Db	183	CGGCCCAAGTCCCGGGCCGAGACATAGTGTGCATGACCGCTGTGGAGGAGATCTCTGAG	242
Qy	165	LeuAlaAlaAspHisProLeuSerArgAsnLysTrpValValGlnLysTyrIle	184
Db	243	CTGGCAGTGCAGACCACTCTTTTCCAGGGACAAAGTGGTGGTCCAGAGTACATC	302
Qy	185	GluThrProLeuLeuIleCysAspThrLysPheAspIleArgGlnTrpPheLeuValThr	204

Db	303	GAGAGCGCGCTGCTCATCTGTGCACCAAGTTTCGACATCAGACAGTGGTTCCTCGTCA	362		
Qy	205	AspTrpAsnProLeuThrIleTrpPheTyrLysGluSerTyrLeuArgPheSerThrGln	224		
Db	363	GACTGGAACCCCTGACCATCTGGTTCTACAAGGAGAGTTACTTGGCGTTCTCAACTCAG	422		
Qy	225	ArgPheSerLeuAspLysLeuAspSerAlaIleHisLeuCysAsnAsnAlaValGlnLys	244		
Db	423	CGTTTCTCCTTGGACAAGCTGGACAGCGCATCCACTGTGCACAAACGCGTCCAGAAG	482		
Qy	245	TyrLeuLysAsnAspValGlyArgSerProLeuLeuProAlaHisAsnMetTrpThrSer	264		
Db	483	TACCTGAAGAATGATGTGGCGCGCAGCCCTGCTGCCCGCACACAACATGTGGACGAGC	542		
Qy	265	ThrArgPheGlnGluTyrLeuGlnArgGlnGlyArgGlyValAlaValTrpGlySerValIle	284		
Db	543	ACCAGGTTCCAGAGTACCTGCAGCGCCAGGCGCGTGGCGCGTGTGGGCGAGCGTCA	602		
Qy	285	TyrProSerMetLysLysAlaIleAlaHisAlaMetLysValAlaGlnAspHisValGlu	304		
Db	603	TACCCGTCCATGAAGAAGGCCATCGCCACGCCATGAAGTGGCCCGACGACGCTGGAG	662		
Qy	305	ProArgLysAsnSerPheGluLeuTyrGlyAlaAspPheValLeuGlyArgAspPheArg	324		
Db	663	CCTCGCAAGAACAGCTTTGAGCTCTACGGGGCTGACTTCGTCCTGTGGANGGACTTCANG	722		
Qy	325	ProTrpLeuIleGluIleAsnSerSerProThrMetHisProSerThr	340		
Db	723	CCCTGGCTGATCGAGATCAATTCAGCCCCACCATGACCCCGTCCAGC	770		
RESULT 7					
AK080321	2520 bp	mRNA	linear	HTC	03-APR-2004
LOCUS	Mus musculus 3 days neonate thymus cDNA, RIKEN full-length enriched				
DEFINITION	library, clone:A630053H17 product:HOITL PROTEIN homolog [Homo				
	sapiens], full insert sequence.				
ACCESSION	AK080321				
VERSION	AK080321.1 GI:26348476				
KEYWORDS	HTC; CAP trapper.				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Mus musculus				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;				
	Sciurognathi; Muridae; Murinae; Mus.				
REFERENCE	1	Carninci, P. and Hayashizaki, Y.			
AUTHORS	High-efficiency full-length cDNA cloning				
TITLE	Meth. Enzymol. 303, 19-44 (1999)				
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)				
PUBMED	11042159				
REFERENCE	2	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,			
AUTHORS	Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.				
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to				
JOURNAL	prepare full-length cDNA libraries for rapid discovery of new genes				
PUBMED	Genome Res. 10 (10), 1617-1630 (2000)				
REFERENCE	3	Carninci, P., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,			
AUTHORS	Konno, H., Akiyama, J., Nishi, K., Kusunagi, T., Tashiro, H., Itoh, M.,				
TITLE	Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,				
JOURNAL	Yamanoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,				
PUBMED	Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M.,				
AUTHORS	Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsumura, S., Kawai, J.,				
TITLE	Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.				
JOURNAL	RIKEN integrated sequence analysis (RISA) system--384-format				
PUBMED	sequencing pipeline with 384 multicapillary sequencer				
AUTHORS	Genome Res. 10 (11), 1757-1771 (2000)				
TITLE	11076861				
JOURNAL	FANTOM Consortium				
PUBMED	Functional annotation of a full-length mouse cDNA collection				
AUTHORS	Nature 409, 685-690 (2001)				

REFERENCE
AUTHORS

5 The PANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.

TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

JOURNAL Nature 420, 563-573 (2002)

REFERENCE 6 (bases 1 to 2520)

AUTHORS Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,T., Kasukawa,T., Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Okazaki,Y., Saito,K., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.

TITLE
JOURNAL

Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-resesgsc.riken.jp, URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)

COMMENT

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Tissues were provided by Dr. John Todd (Dept. of Medical Genetics Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome Trust/WRC building Addenbrookes Hospital Cambridge) whose assistance we gratefully acknowledge. Please visit our web site for further details.

URL:http://genome.gsc.riken.jp/
URL:http://fantom.gsc.riken.jp/

FEATURES

source

```
1. .2520
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:A630053H17"
/db_xref="taxon:10090"
/clone="A630053H17"
/tissue_type="thymus"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stages="3 days neonate"
84. .2198
/note="unnamed protein product; HOTT1L PROTEIN homolog [Homo sapiens] (SPTRIQ9UI99, evidence: FASTY, 85.8%ID, 93.4%length, match=821)
putative"
/codon_start=1
/protein_id="BAC37878.1"
/db_xref="GI:26348477"
```

CDS

```
translation="WDTSSPFVAAGQGEASPPSLPVSTISSPSLQSRMVRNETPYLIWT
TERVDLCRFLSKDMINHYARAGFTTKVGLCLNRLNLPWFDEADADSPFRCLVLG
AEDDKATIEDFLTAARLVKLVVKKSEKQSISIQAREBEAEDTPQPKOEKVLVT
VSSDFDVALSCQBHLSIAHOKIDKPNPSFLYSPDDWSQFLQRYQIVHEGAELR
YLEVQVQCEDILOQLQNVVQLDMDGDRNIWVKGAKSRGRGIMCNRLDMLKLV
DCNPLMKDGMWIKVYIERPLIFGTFKFLRWPLVTDWNLPTWVFRDVSIRFSTQ
PFLKNLNSVHLNNSIQHLEASCRHPMLPDPNMWSSORFQNLHLEVDAPKAWSS
VIVFQMKAAVTHALQTSQDNQVCKASFELEGADVFQFQPMLEINAFSTWAPST
AVTARLCAGVQADTLRVVDRRLDKSCDTGAPFLYKQPAVEVQYVIRLLVEGSTI
KKPVPHRRRTGVSRLPHLLTQQGSGSKDSGSPHRSASRKNARASLEHTEKPEP
AAVASVSGKGPAPFPPLSHKAWLPSPVRHPQGRVLRHQDLQVGSKALSTTGA
LMTPLTAKVLSPFPHLDLKLAPSLMKPGKVGFEFLCCCTTWRVVLSSGIGEGHQRRA
PRPSSAPGKGLSSTFPCSKTET"
```

ORIGIN

Alignment Scores: 1.37e-80 Length: 2520
Pred. No.: 1052.00 Matches: 226
Score: 1052.00
Percent Similarity: 55.1% Conservative: 85
Best Local Similarity: 40.1% Mismatches: 177
Query Match: 36.9% Indels: 76
DB: 4 Gaps: 8

US-10-635-977-2 (1-541) x AK080321 (1-2520)

QY 2 AlaSerSerIleLeuLysTrpValSerHisGlnSerCysSerArgSerArgSer 21
438 GCCCGCAAGCTCTCAAGCTGTGTGAAGTGGAGAGAGAGTCAAGTCCATCTCCATC 497
QY 22 LysProArgAspGlnArgGluGluAlaGlySerSerAspLeuSerArgGlnAspAla 41
498 CAGGCAAGAGAG-----GAAGAGGCCCCAGAGAGATACACAGCCCCAGAAACAG----- 545
QY 42 GluAsnAlaGluAlaLysLeuArgGlyLeuProGlyGlnLeuValAspIleAlaCysLys 61
546 -----GAAAGAAAGCTAGTACCGTGTCTCAGACTTTGTGCATGAGGCTCTGAGT 596
QY 62 ValCysGlnAlaTyrluLeuGlnLeuGluHisGluAspIleAspThrSerAlaAspAla 81
597 GCATGCCAGGAGCCTTTAGCAGCATAGCCCCAAGACATTTGACAAAGACCCCCAACCTC 656
QY 82 ValGluAspLeuThrGluAlaGluTrpGluAspLeuThrGlnGlnTyrluSerLeuVal 101
657 CGCTCTACCTCAGCCCCGATGTTGGTCCCAAGTTCCTTACGGCTTACCAATAGTT 716
QY 102 HisGlyAspAlaPheIleSerAsnSerArgAsnTyrluPheSerGlnCysGlnAlaLeu 121
717 CATGAAGGGCAGAGCTCAGATACCTGGAGGTCCAAAGTCCAGCGCTGTGAAGACATCTTA 776
QY 122 AsnArgIleThrSerValAsnProGlnThrAspIleAspGlyLeuArgAsnIleTrpIle 141
777 CAGCAGCTTCAGACGTGTGTTACCCAGTTAGACATGAGGGGGGATCGGAACATCTGGATC 836
QY 142 IleLysProAlaAlaLysSerArgGlyArgAspIleValCysMetAspArgValGluLys 161
837 GTGAAGCCCGGAGCCNAGTCCGAGCGGAGGATTTGTGATGATGAACCGCTGGATGAG 896
QY 162 IleLeuGluLeuAlaAlaAlaAspHisProLeuSerArgAspAsnLysTrpValValGln 181
897 ATGCTGAAGCTGGTGGAGCTGCACACCCCATGCTCATGAAGGATGGCAAGTGGATCGTCAG 956
QY 182 LysTyrlleGluThrProLeuLeuIleCysAspThrLysPheAspIleArgGlnTrpPhe 201
957 AAGTACATTGAGCGGCCCTCTTATCTTCGGCACCAAGTTTGACCTGAGACAGTGGTTC 1016
QY 202 LeuValThrAspTrpAsnProLeuThrIleTrpPheTyrluLysGluSerTyrluArgPhe 221
1017 CTGTGACTGACTGGNACCCACTACCGTGTGGTGTCTACCGAGACAGCTACATTCGTTTC 1076
QY 222 SerThrGlnArgPheSerLeuAspLysLeuAspSerAlaIleHisLeuCysAsnAla 241
1077 TCCACACAGCCCTTCTCCCTGAAGAACCTCGACAACTCTGTGCACCTGTGTAAACAATCC 1136
QY 242 ValGlnLysTyrluLeuLysAsnAspValGlyArgSerProLeuLeuProAlaHisAsnMet 261
1137 ATCCAGAGACACTTGGAGGCTTCTGTGTCACCGGACCCCGATGCTGCCCCAGATAACATG 1196
QY 262 TrpThrSerThrArgPheGlnGluTrpLeuGlnArgGlnGlyArgGlyAlaValTrpGly 281
1197 TGTTCACGACAGAGTTCAGGCCCATTTGAGGAGGTAGATGCCCAAGAGCCCTGGTCC 1256
QY 282 SerValIleTyrluProSerMetLysLysAlaIleAlaHisAlaMetLysValAlaGlnAsp 301
1257 AGCGTCATCGTCCAGGATGAAGGCTGCTGTGTATCCATCCCTCGACAGCTCCCAAGAT 1316
QY 302 HisValGluProArgLysAsnSerPheGluLeuTyrluGlyAlaAspPheValLeuGlyArg 321
1317 AACGTGAGTCCCGAAGGCCAGCTTTGAGCTCTATGGGCGAGACTTTGTGTTGGGAA 1376

Qy	322	AspPheArgProTrpIleuLeuGluLeuAanSerSerProThrMetHisProSerThrPro	341
Db	1377	GACTTCAGCCCTGGTGAATGAATCAATGCGACCCACACCTGCACATGCCAGGT	1436
Qy	342	ValThrAlaGlnLeuCysAlaGlnValGlnGluAspThrIleLysValAlaVal-----	359
Db	1437	GTCACTGCCCGCTCTGTCCCGTGTGCAGCGAGATACCTGGGTGTGTCATCGACCGG	1496
Qy	360	-----AspArgSerCysAspIleGlyAanPheGluLeuLeuTrpArgGlnProValVal	377
Db	1497	CGCTGGACCGCAGCTGTGATACCGGAGCCTTTGAGCTCATCTATAAGCAGCCTGCTGTG	1556
Qy	378	GluProProProPheSerGlySerAspLeuCysValAlaGlyValSerValArgAla	397
Db	1557	GAGGTGCTCAGTACGTGGGTATCCGGCTCTAGTGGAGGGCTTACCATCAAGAAAG---	1613
Qy	398	ArgArgGlnValLeuProValCysAanLeuLysAlaSerAlaSerLeuLeuAspAlaGln	417
Db	1613	-----	1613
Qy	418	ProLeuLysAlaArgGlyProSerAlaMetProAspProAlaGlnGlyProProSerPro	437
Db	1614	-----CCCGTGCCC	1622
Qy	438	AlaLeuGlnArgAspLeuGlyLeuLysGluGluLysGlyLeuProLeuAlaLeuLeuAla	457
Db	1623	GTGGGTGTCATCGGGGACAGGGGTCCGCTCATCA-----CTCCCTCATCTGCTG	1670
Qy	458	ProLeuArgGlyAlaAlaGluSerGlyGlyAlaAlaGlnProThrArgThrLysAlaAla	477
Db	1671	ACCAGCAAGGCTCTGGGAAAGCAAGGATCAGGTCCTCCCTACCCAC---AGTCAGCT	1727
Qy	478	GlyLysValGluLeuProAlaCysProCysArgHisValAspSerGlnAlaProAanThr	497
Db	1728	TCTAGGAAATGCTAGGGCCGAGAGCTGGAGCACACCGAGAAGCGGAACCTGGGCC	1787
Qy	498	GlyValProValAlaGlnPro-----	505
Db	1788	GTGCGCTCTGCTCCGGAAGGGGAGAAAGCCCTTTCCACTTCCCTACGCTCCATTC	1847
Qy	506	LysSerTrpAspProAanGlnLeuAanAlaHisProLeuGluProValLeuArg-----	523
Db	1848	AAGCCTGTGCTGCTCTCCCGGTGCACCGACCGCCGCGGTCTCAGACTCCAA	1907
Qy	524	-----GlyLeuLysThrAlaGluGlyAlaLeuArgProPro	535
Db	1908	CATGACCAGCTGGTGGCTTAAGGCTCTGTCAACACAGCAAGGCTTTGATGACTCTT	1967
Qy	536	ProGlyGlyLys	539
Db	1968	CCTACTGCCAAG	1979
RESULT 8			
HSM808322			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
CONSTRM			
TITLE			
JOURNAL			
Submitted (22-SEP-2004) MIPs, Ingolstaedter Landstr.1, D-85764			
Neueberg, GERMANY			
COMMENT			
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by EMBL (European Molecular Biology Laboratories, Heidelberg/Germany) within the cDNA sequencing consortium of the German Genome Project.			
This clone (DKFZp686D076) is available at the RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH in Berlin, Germany. Please contact RZPD for ordering:			
http://www.rzpd.de/cgi-bin/products/cl.cgi?cloneID=DKFZp686D076			
Further information about the clone and the sequencing project is available at http://mips.gsf.de/projects/cdna/.			
FEATURES			
source			
1. .4184			
/organism="Homo sapiens"			
/mol_type="mRNA"			
/db_xref="RZPD:DKFZp686D076Q"			
/db_xref="taxon:9606"			
/clone="DKFZp686D076"			
/issue_type="cervix"			
/clone_lib="686 (synonym: hlcc3). Vector pSport1_Sfi; host DH10B; sites SfiIA + SfiIB"			
/dev_stage="adult"			
/note="hypothetical protein"			
1. .4184			
/gene="DKFZp686D076"			
2529. .3587			
/gene="DKFZp686D076"			
/codon_start=1			
/product="hypothetical protein"			
/protein_id="CAH10554.1"			
/db_xref="GI:50949619"			
/translation="MDIDKDLAPLYLTPEGWSLFLQRYVQVVEGAELRHLDTVQR			
CEDILQQLAAVVPQIDMEGRNITWIKPKAKSRGIMCMMDHLEMLKLNVHNPVVMK			
DKGVVQKTIERPLIFGKFDLRWFLVTDWNPPLTWFRDSYIRFSQPPSLKLD			
NSVHLCNNSIQHLENSCHRPPLPQNNSSQRPQALHQEGAPNAWSTIIVPGMKD			
AVTHALQTSQTVQCRKASFELYGADVFEGEDFQPLIIBINASPTWAPSTAVTARLCA			
GVQADTLRVIVDMLDRNCDTGAFELIYKQPVTTSPASTPGPSLLEPMYSIDTRRSSD			
DSTASWALRPCRQARP"			
ORIGIN			
Alignment Scores:			
Pred. No.: 1-82e-77 Length: 4184			
Score: 1020.00 Matches: 220			
Percent Similarity: 55.5% Conservative: 69			
Best Local Similarity: 42.2% Mismatches: 151			
Query Match: 35.7% Indels: 81			
DB: 4 Gaps: 12			
US-10-635-977-2 (1-541) x HSM808322 (1-4184)			
Qy	54	GlnLeuValAspIleAlaCysLysValCysGlnAlaTyrlLeuGlyGlnLeuGluHisGlu	73
Db	2472	GAGTTTGTGTAAGAGCTCTGTGTGCGTGGAGAGTACCTTAGCAACTTGCCCATG	2531
Qy	74	AspIleAspThrSerAlaAspAlaValGluAspLeuThrGluAlaGluTrpGluAspLeu	93
Db	2532	GACATCGACAGACCTGGAGGCCCGCTGTACTCACCCCGAGGGCTGGTCCCTCTTC	2591
Qy	94	ThrGlnGlnTyrlTyrlSerLeuValHisGlyAspAlaPheIleSerAanSerArgAsnTyrl	113
Db	2592	CTCCAGCGCTACTACCAAGTGTGTCCAGAGGGGCGAGAACTCGACACTCAG	2651
Qy	114	PhaSerGlnCysGlnAlaLeuLeuAanArgIleThrSerValAanProGlnThrAspIle	133
Db	2652	GTCCAGCGCTGTGAGGACATCTCGACGACCTGAGCGCGGTGTACCCAGATAGACATG	2711
Qy	134	AspGlyLeuArgAanIleTrpIleIleLysProAlaAlaLysSerArgGlyArgAspIle	153
Db	2712	GAGGGGATCGCAACATCTGGATCGTGAAGCCAGGACCAAGTCCCGCGACGAGGCATC	2771
Qy	154	ValCysMetAspArgValGluGluLeuLeuGluLeuAlaAlaAspHisProLeuSer	173
Db	2772	ATGTGCATGACCACTGGAGGAGATGCTGAAGCTGTGAACCGCAACCCCGTGTGATG	2831


```
QY 217 SerTyrLeuArgPheSerThrGlnArgPheSerLeuAspSerAlaLeuHis 236
Db 64 AGTACCTCGCTTCTCCACACGCGCTTCTCCGACCAAGCTGACGCGCATCCAC 123
QY 237 LeuCysAsnAlaValGlnLysTyrLeuLysAsnAspValGlyArgSerProLeuLeu 256
Db 124 CTGTGTAAACAACCTCCATCCAGAGCGCTCAAAATGATAAGAGCGCAGTCCGCTGCTA 183
QY 257 ProAlaHisAsnMetTyrThrSerThrArgPheGlnGluTyrLeuGlnArgGlnGlyArg 276
Db 184 CCTTGTTCATAACATGTGGACCACTCGTTTCCAGAGTACCTGCGAGAAGAGGCGCGCA 243
QY 277 GlyAlaValTyrGlySerValLleTyrProSerMetLysLysAlaLeuHisAlaMet 296
Db 244 GGAGGAACGTGGGTGAGCATCATCTACCCGCTCTATGAAAAGAGCTGTCAACATGCCATG 303
QY 297 LysValAlaGlnAspHisValGluProArgLysAsnSerPheGluLeuTyrGlyAlaAsp 316
Db 304 AGGGTAGCCCGAGGACCACTAGAACGCCGTAAAGACAGCTTTCAGGAGCGGAC 363
QY 317 PheValLeuGlyArgAspPheArgProTyrLeuLleGluLleAsnSerSerProThrMet 336
Db 364 TTCATCTGNGCGGAGACTTCAAGCCATGCTTATCGAGATCAACTCCAGCCGCCCATG 423
QY 337 HisProSerThrProValThrAlaGlnLeuLysAlaGlnValGlnGluAspThrLleLys 356
Db 424 CACCCCTCCACTCCCGTACAGCTCAGCTCTGTGCCAGCGGTGAGGAGGACCATCAAG 483
QY 357 ValAlaVal-----AspArgSerCysAspIleGlyAsnPheGluLeuLeuTyr 372
Db 484 GTGTGTGTGGATCGCAAACTGGACCGAACTGTGACATCGGCACTTTGAGCTTCTGTGG 543
QY 373 ArgGlnProValValGluProProPheSerGlySerAspLeuLysValAlaGlyVal 392
Db 544 CGGAGCTCGCGTGGAGCTGCCACCGCTTCAACGGGTCTGACCTCTGTGTGGAGGTATC 603
QY 393 SerValArgAlaArgArgGlnValLeuProValCysAsnLeuLysAlaSerAlaSer 412
Db 604 AGCGTGAAGAAAGCAAGAACAAATGCCGCCATTCGCCAGCGTTCGCCCTCTCAGAATCA 663
QY 413 LeuLeuAspAlaGlnProLeuLysAlaArgLysProSerAlaMetProAspProAlaGln 432
Db 664 CTCTTAGATGCTCCACCAGTGGCGAGTG-CCCCGGCGCTGTATGGAAACAGTGATC 722
QY 433 GlyProProSerProAlaLeuGlnArgAspLeuLysGluGlu 448
Db 723 AGGCTTCCAGACGACCGCTGGCGGAGGACTGGGAACGGGAAGAA 770

RESULT 10
DQ045868
LOCUS
DEFINITION
1059 bp DNA linear GSS 02-JUN-2005
Homo sapiens ARPC4 gene, VIRTUAL TRANSCRIPT, partial sequence,
genomic survey sequence.
ACCESSION
DQ045868
VERSION
DQ045868.1 GI:66897083
KEYWORDS
GSS.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1 (bases 1 to 1059)
Nielsen,R., Bustamante,C., Clark,A.G., Glanowski,S., Sackton,T.B.,
Hubisz,M.J., Fedel-Alon,A., Tanenbaum,D.M., Civallo,D.,
White,T.J., Sninsky,J.J., Adams,M.D. and Cargill,M.
A Scan for Positively Selected Genes in the Genomes of Humans and
Chimpanzees
(ef) PLoS Biol. 3 (6), E170 (2005)
JOURNAL
PUBMED
15869325
REFERENCE
2 (bases 1 to 1059)
Nielsen,R., Bustamante,C., Clark,A.G., Glanowski,S., Sackton,T.B.,
Hubisz,M.J., Fedel-Alon,A., Tanenbaum,D.M., Civallo,D.,
```

```
White,T.J., Sninsky,J.J., Adams,M.D. and Cargill,M.
Direct Submission
Submitted (05-MAY-2005) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT
This sequence was made by sequencing genomic exons and ordering
them based on alignment. Translation starts at the beginning of
alignment.
FEATURES
source
1..1059
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="3"
<1..>1059
/gene="ARPC4"
/locus_tag="HC11626"
ORIGIN
Alignment Scores:
Pred. No.: 1,46e-68 Length: 1059
Score: 908.00 Matches: 165
Percent Similarity: 73.2% Conservatives: 48
Best Local Similarity: 56.7% Mismatches: 78
Query Match: 31.8% Indels: 0
DB: 11 Gaps: 0
US-10-635-977-2 (1-541) x DQ045868 (1-1059)
QY 74 AspIleAspThrSerAlaAspAlaValGluAspLeuThrGluAlaGluTrpGluAspLeu 93
Db 4 GACATCGACAAGACCTGGAGGCCCGCTGACTCACCCCGGAGGCTGGTCCCTCTTC 63
QY 94 ThrGlnGlnTyrTyrSerLeuValHisGlyAspAlaPheIleSerAsnSerArgAsnTyr 113
Db 64 CTCACGCGTACTACCAAGTGGTCCAGAGGGCGAGAACTCAGGCACCTCGACACTCAG 123
QY 114 PheSerGlnCysGlnAlaLeuAsnArgIleThrSerValAsnProGlnThrAspIle 133
Db 124 GTCCAGCGCTGTGAGACATCTTCGACAGCTGCAGCGCGTGTACCCAGATAGACATG 183
QY 134 AspGlyLeuArgAsnIleTrpIleLysProAlaAlaLysSerArgGlyArgAspIle 153
Db 184 GAAGGGGATCGCAACATCTGGATCGTGAAGCCAGGAGCAAGTCCGCGGACGAGGCATC 243
QY 154 ValCysMetAspArgValGluGluLleLeuGluLeuAlaAlaAspHisProLeuSer 173
Db 244 ATGTGATGACCACTCGGAGGAGATGCTGAAGCTGGTGAACCGCAACCCCGGTGGTATG 303
QY 174 AtgAspAsnLysTyrValGlnLysTyrIleGluThrProLeuLleLysCysAspThr 193
Db 304 AAGACGGCAAGTGGGTGGTGGAGAGTATATTGAGCGGCCCTCTCTCATCTTTGGCACC 363
QY 194 LysPheAspIleArgGlnTrpPheLeuValThrAspTrpAsnProLeuThrIleTrpPhe 213
Db 364 AAGTTTGACCTCAGACAGTGGTTCCTGGTAACTGACTGGAACCCACTTACCGTGTGTTTC 423
QY 214 TyrLysGluSerTyrLeuArgPheSerThrGlnArgPheSerLeuAspLysLeuAspSer 233
Db 424 TACCGGACAGCTATATTCGCTTTTCCAGCAGCCCTTCTCCCTGAAGAACCTCGGACAC 483
QY 234 AlaIleHisLeuCysAsnAsnAlaValGlnLysTyrLeuLysAsnAspValGlyArgSer 253
Db 484 TCAGTGACCTGTGCAACAACTCCATCCAGAGCACTCGAGAACTCATCCCATCGGCAT 543
QY 254 ProLeuLeuProAlaHisAsnMetTyrThrSerThrArgPheGlnGluTyrLeuGlnArg 273
Db 544 CCATGCTTCGCGCAGACAACATGTGCTAGCCAGAGGTTCCAGGCGCCACCTCGAGAG 603
QY 274 GlnGlyArgGlyAlaValTyrGlySerValIleTyrProSerMetLysLysAlaIleAla 293
Db 604 ATGGGTGCCCAATGCTTGGTCCACCATCATCGTGCCTGGCATGGAAGGATGTGTGATC 663
QY 294 HisAlaMetLysValAlaGlnAspHisValGluProArgLysAsnSerPheGluLeuTyr 313
```

```

Db      664  CAGCAGCTTCAGACCTCCAGGACACCGTCAGTGTGGAGGCCAGCTTTGAGCTCTAT 723
Qy      314  GlyAlaAspPheValLeuGlyArgAspPheArgProTrpLeuIleGluIleAanSerSer 333
Db      724  GCGCTGACTTCGTGTGGGGAGGACTTCAGCCCTGGCTGATTGAGATCAACGCCAGC 783
Qy      334  ProThrMetHisProSerThrProValThrAlaGlnLeuCysAlaGlnValGlnGluAsp 353
Db      784  CCACAGATGACACCTCCACAGCAGTCACTGCCCGGCTGTGCTGGCGTGAAGCTGAC 843
Qy      354  ThrIleLysValAlaValAspArgSerCysAsp 364
Db      844  ACCCTGGCGTGTCAATTGACCGATGCTGGAC 876

```

RESULT 11

```

BM808637 1023 bp mRNA linear EST 05-MAR-2002
LOCUS AGENCOURT_6617786 NIH_MGC_124 Homo sapiens cDNA clone IMAGE:5734232
DEFINITION 5', mRNA sequence.

```

```

ACCESSION BM808637
VERSION 1 GI:19125460
KEYWORDS EST.
SOURCE Homo sapiens (human)

```

ORGANISM

```

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.

```

```

1 (bases 1 to 1023)

```

REFERENCE

```

AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)

```

COMMENT

```

Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Invitrogen
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone Distribution: Agencourt Bioscience Corporation
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov

```

```

Plate: LLAM12739 row: d column: 09

```

```

High quality sequence start: 6

```

```

High quality sequence stop: 664.

```

FEATURES

source

```

1..1023
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5734232"
/tissue_type="hippocampus"
/lab_host="DH10B"
/clone_lib="NIH MGC 124"
(note="Organ: brain; Vector: pCMV-SPORT6; Site 1: EcoRV
(destroyed); Site 2: NotI; RNA source male hippocampus,
age 27. Library is oligo-dT primed and directionally
cloned (EcoRV site is destroyed upon cloning). Average
insert size 1.4 kb, insert size range 0.9-4 kb. Library is
normalized and enriched for full-length clones and was
constructed by C. Gruber (Invitrogen). Research Genetics
tracking code 012.")

```

ORIGIN

```

Alignment Scores: 4.23e-67 Length: 1023
Pred. No.: 891.00 Matches: 177
Score: 67.7% Conservative: 54
Percent Similarity: 51.9% Mismatches: 94
Best Local Similarity: 31.2% Indels: 16
Query Match: 3 Gaps: 4
DB:

RESULT 12
BM808516 1022 bp mRNA linear EST 05-MAR-2002
LOCUS AGENCOURT_6617783 NIH_MGC_124 Homo sapiens cDNA clone IMAGE:5734160
DEFINITION 5', mRNA sequence.

```

```

107 IleSerAsnSerArgAsnTyrPheSerGln-----CysGlnAlaLeuLeuAanArg 123
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
14  GTCCCGCGCAGCAGGACGACCTCTACACTCAGGTCCAGCGTGTGAGGACATCTCTGCAGCAG 73
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
124 IleThrSerValAsnProGlnThrAspIleAspGlyLeuArgAsnIleTrpIleLys 143
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
74  CTGAGCGCTGTGTACCCAGATAGACATGGAAGGGGATCGCAACATCTGGATCGTGAAG 133
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
144 ProAlaAlaLysSerArgGlyArgAspIleValCysMetAspArgValGluGluLeu 163
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
134  CCAGGAGCCAAAGTCCCGCGGACGAGCATCATGTGATGACCACTGAGGAGATGCTG 193
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
164 GluLeuAlaAlaAspHisProLeuSerArgAspAsnLysTrpValValGlnLysTrp 183
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
194  AAGCTGCTGAACGGCAACCCCGTGTGTATGAAGGACGCAAGTGGGTGTCGACAGAATAT 253
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
184  IleGluThrProLeuLeuIleCysAspThrLysPheAspIleArgGlnTrpPheLeuVal 203
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
254  ATTGAGCGGCCCTCTCTCATCTTTGGCACCAAGTTGACCTCAGACAGTGGTTCTCGGTA 313
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
204  ThrAspTrpAsnProLeuThrIleTrpPheTrpLysGluSerTyrLeuArgPheSerThr 223
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
314  ACTGACTGGAACCCACTTACCGTGTGGTCTTACCGGACAGCTATATCCGCTTTTCACG 373
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
224  GlnArgPheSerLeuAspLysLeuAspSerAlaIleHisLeuCysAsnAsnAlaValGln 243
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
374  CAGCCCTTCTCCCTGAAGAACCTGGCAACTCAGTGCACCTGTGCAACAACATCCATCCAG 433
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
244  LysTrpLeuLysAsnAspValGlyArgSerProLeuLeuProAlaHisAsnMetTrpThr 263
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
434  AAGCACTCGGAGAACTCATGCCATCGGCATCCACATGCTTCCGCCAGACAACTATGGTCT 493
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
264  SerThrArgPheGlnGluTyrLeuGlnArgGlnGlyArgGlyAlaValTrpGlySerVal 283
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
494  AGCCAGAGTTCCAGGCCACCTCGCAGGAGATGGGTGCCCAAAATGCTTGGTCCCACTTC 553
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
284  IleTyrProSerMetLysLysAlaIleAlaHisAlaMetLysValAlaGlnAspHisVal 303
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
554  ATCGTGTCTGGCATGAAGATGCTGTGATCCACGACATTCAGACCTCCCGCAGCACCCG 613
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
304  GluProArgLysAsnSerPheGluLeuTyrGlyAlaAspPheValLeuGlyArgAspPhe 323
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
614  CAGTGTGGAAGGCCAGCTTTGAGCTCTATGCGCTGACTTCTGTCGCGGAGGACTTC 673
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
324  ArgProTrpLeuIleGluIleAsnSerSerProThrMetHisProSerThrProValThr 343
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
674  CAGCCCTGCTGATTGAGATCAACGCCAGCCCACTGCGCACCTTCCACACGACGACTCT 733
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
344  AlaGlnLeuCysAlaGlnValGlnGluAspThrIleLysValAlaValAspArg----- 361
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
734  GCCCGGCTCTGTGCTGGCGTGCAAGCTGACACCCCTGCGTGTCTATTGACCGGATGCTG 793
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
362  -----SerCysAspIleGly-AsnPheGluLeuLeuTrpArgGlnProVal---ValGl 378
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
794  GACCGGCACTGTGACCCAGGAGCCCGTTGAGCTCATCTATAAGCAGCCCGTAACCACTTC 853
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
378  uPro-ProProPheSerGlySerAspLeuCysValAlaGlyValSerValArgAlaAa 398
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
854  CCCAGCTTCCCAAGGCCCGGCTGCTTTCTCCCATGTACTCCGAAACACACAGGGG 913
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
398  rGArgGlnValLeuProValCysAsnLeuLysAlaSerAlaSerLeuLeuAspAlaGlnP 418
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
914  CCAGG-----TCCCTCAGAACGACGACCCACAGCAGGCTGGTGGGGG 955
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
418  roLeuLysAlaArgGlyProSerAlaMetProAspProAlaGlnGlyProProSerPro 437
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
956  AACTAAGTGCCTTGGTCGACCCACTGGGCAAGGGCCCTTGTAGGGGACTCTCTTACCCA 1014
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::

```

BM808516
 VERSION BM808516.1 GI:19125339
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Homnidae; Homo.
 REFERENCE 1 (bases 1 to 1022)
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Invitrogen
 CDNA Library Preparation: Life Technologies, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLAM12739 row: a column: 09
 High quality sequence start: 78
 High quality sequence stop: 658.
 FEATURES
 source
 1..1022
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:5734160"
 /tissue_type="hippocampus"
 /lab_host="DH10B"
 /clone_lib="NIH_MGC_124"
 /note="Organ: brain; Vector: pCMV-SPORT6; Site 1: EcoRV
 (destroyed); Site 2: NotI; RNA source male hippocampus,
 age 27. Library is oligo-dT primed and directionally
 cloned (EcoRV site is destroyed upon cloning). Average
 insert size 1.4 kb, insert size range 0.9-4 kb. Library is
 normalized and enriched for full-length clones and was
 constructed by C. Gruber (Invitrogen). Research Genetics
 tracking code 012."
 ORIGIN
 Alignment Scores:
 Pred. No.: 4,13e-60 Length: 1022
 Score: 811.00 Matches: 162
 Percent Similarity: 64.2% Conservative: 48
 Best Local Similarity: 49.5% Mismatches: 72
 Query Match: 28.4% Indels: 46
 Db: 3 Gaps: 5
 US-10-635-977-2 (1-541) x BM808516 (1-1022)
 QY 116 GlnCysGlnAlaLeuLeuAsnArgIleThrSerValAsnProGlnThrAspIleAspGly 135
 Db 41 CGCTGTGAGGACATCTGCAGCAGCTGCAGGCGGTGTATCCACATAGACATGGAAAGG 100
 QY 136 LeuArgAsnIleTrrpIleIleLysProAlaLalysSerArgGlyArgAspIleValCys 155
 Db 101 GATCGCAACATCTGGATCOTGAAGCCAGAGCCAAAGTCCGCGGACGAGGCATCATGTGC 160
 QY 156 MetAspArgValGluGluIleuGluLeuAlaAlaAAspHisProLeuSerArgAsp 175
 Db 161 ATGGACACCACTTGAGGAGATGCTGAAGCTGTGTGAACCGCAACCCCGTGTGTATGAAGGAC 220
 QY 176 AsnLysTrpValValcIlnLysTrIleGluThrProLeuLeuIleCysAspTrpLysPhe 195
 Db 221 GGCAAGTGGTGGTGCAGAGATATATTGAGCGGCCCTCTCATCTTTGGCACCAAGTTT 280
 QY 196 AspIleArgGlnTrpPheLeuValThrAspTrpAsnProLeuThrIleTrpPheTrpLys 215
 Db 281 GACCTCTTACAGTGGTCTCTGTAACATGACATGGAAACCACTTACCGTGTGGTTCTACCGC 340

TITLE Direct Submission
 JOURNAL Submitted (05-MAY-2005) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA
 COMMENT This sequence was made by sequencing genomic exons and ordering them based on alignment. Translation starts at the beginning of alignment.

FEATURES

source

1..1059
 /organism="Pan troglodytes"
 /mol_type="genomic DNA"
 /db_xref="taxon:9598"
 <1..>1059
 /gene="ARPC4"
 /locus_tag="HC11626"

ORIGIN

Alignment Scores:
 Pred. No.: 4,35e-60 Length: 1059
 Score: 811.00 Matches: 154
 Percent Similarity: 67.0% Conservative: 41
 Best Local Similarity: 52.9% Mismatches: 96
 Query Match: 28.4% Indels: 0
 DB: 11 Gaps: 0

US-10-635-977-2 (1-541) x DQ045869 (1-1059)

QY 74 AspIleAspThrSerAlaAspAlaValGluAspLeuThrGluAlaGluTrpGluAspLeu 93
 DB 4 GACATCGACAGGACCTGGAGGCCCTACCTGTACCTACCCAGGCGTGGTCCCTCTTC 63
 QY 94 ThrGlnGlnTyrSerLeuValHisGlyAspAlaPheIleSerAsnSerArgAsnTyr 113
 DB 64 CTCACGCGCTACTACCAAGTGTCTCCAGGAGGGCAGAACTCAGGCACCTCCACTCAG 123
 QY 114 PheSerGlnGlnAlaLeuAsnArgIleThrSerValAsnProGlnThrAspIle 133
 DB 124 GTCCAGCGCTGTGAGACATCTCTGACGAGCTGCGGGCGTGGTACCCAGATACATG 183
 QY 134 AspGlyLeuArgAsnIleTrpIleLeuPheProAlaAlaLysSerArgGlyArgAspIle 153
 DB 184 GAAGGGGATCGCAACATCTGGATCTGGAAGCAGGAGCAAGTCCCGGGAGGAGGCATC 243
 QY 154 ValCysMetAspArgValGluLeuLeuGluLeuAlaAlaAspHisProLeuSer 173
 DB 244 ATGTGCATGGACCACTCGAGAGATGCTGAAGCTGTGAACGCAACCCCGTGGTATG 303
 QY 174 ArgAspAsnLysTrpValGlnLysTyrIleGluThrProLeuLeuIleCysAspThr 193
 DB 304 AAGGACGGCAAGTGGTGGTGAGAGATATATTGAGCGGCGCCCTCTCATCTTTGGCAC 363
 QY 194 LysPheAspIleArgGlnTrpPheLeuValThrAspTrpAsnProLeuThrIleTrpPhe 213
 DB 364 AAGTTTGACCTCAGACAGTGTCTCTGTAACTGACTGGACCCACTTACCGTGGTTC 423
 QY 214 TyrLysGluSerTyrLeuArgPheSerThrGlnArgPheSerLeuAspLysLeuAspSer 233
 DB 424 TACCGCGACAGTATATCGCTTTTCCAGCGAGCCCTCTCCCTGAAGAACTGGACAAN 483
 QY 234 AlalleHisLeuCysAsnAsnAlaValGlnLysTyrLeuLysAsnAspValGlyArgSer 253
 DB 484 NNN 543
 QY 254 ProLeuLeuProAlaHisAsnMetTrpThrSerThrArgPheGlnGluTyrLeuGlnArg 273
 DB 544 CCATGTCTTCGCCAGACAAATGNNNTCTAGCCGAGGTTCAGGCCACCTGCAGGAG 603
 QY 274 GlnGlyArgGlyAlaValTrpGlySerValIleTyrProSerMetLysLysAlaIleAla 293
 DB 604 ATGGGTGCCCAAAATGCTTGGTCCACCATCATCTGTCGCTGAGGATGCTGTGATC 663
 QY 294 HisAlaMetLysValAlaGlnAspHisValGluProArgLysAsnSerPheGluLeuTyr 313
 DB 664 CACGCACTTACAGCTCCAGGACACCGTGCAGTGTGGAAGGCGCAGCTTTGAGCTCTAT 723

QY 314 GlyAlaAspPheValLeuGlyArgAspPheArgProTrpLeuIleGluLeuSerSer 333
 DB 724 GCGCGTGACTTCGTGTGGGGAGGACTTCCAGCCCTGGCTGATTGAGATCAAGCCAGC 783
 QY 334 ProThrMetHisProSerThrProValThrAlaGlnLeuCysAlaGlnValGlnGluAsp 353
 DB 784 CCCAGATGGCACCTCCACAGCAGTCACCTGCGGCTCTGTGCTGGCGTGAAGCTGAC 843
 QY 354 ThrIleLysValAlaValAspArgSerCysAsp 364
 DB 844 ACCCTGGCGTGTCTATTGACCGGATGCTGGAC 876

RESULT 14

CK597795 777 bp mRNA linear EST 22-JAN-2004
 LOCUS AGENCOURT 17837841 NIH_MGC_238 Rattus norvegicus cDNA clone
 DEFINITION IMAGE:713519 5', mRNA sequence.

CK597795

CK597795.1 GI:41110910

EST.

Rattus norvegicus (Norway rat)

ORGANISM

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muridea; Muridae; Murinae; Rattus.

NIH-MGC http://mgi.nci.nih.gov/

1 (bases 1 to 777)

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Daniela S. Gerhardt, Ph.D.

Office of Cancer Genomics

National Cancer Institute / NIH

Bldg. 31 Rm10A07 Bethesda, MD 20892

Email: cgabs-remail.nih.gov

Tissue Procurement: Howard Jacobs

cDNA Library Preparation: Express Genomics

DNA Sequencing Arrayed by: The I.M.A.G.E. Consortium (LLNL)

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAM15042 row: 0 column: 05

High quality sequence stop: 693.

Location/Qualifiers

1..777

/organism="Rattus norvegicus"

/mol_type="mRNA"

/db_xref="taxon:10116"

/clone="IMAGE:713519"

/tissue_type="testis, pooled"

/lab_host="DH10B TON"

/clone_lib="NIH_MGC_238"

/note="Organ: testis; Vector: pExpress-1; Site: 1: EcoRV;

Site 2: NotI; RNA obtained from testis tissue of 8 wk old

animal. Tissues were snap-frozen and kept at -80C before

RNA extraction and purification (Tri-reagent method). cDNA

was primed using oligo-dT primer:

5'-pGATAGTTCATGATCGGAGCGGCCCTTC-3' and cloned into

the EcoRV/NotI sites of pExpress-1. Size-selection >1.4kb

resulted in an average insert size of 1.9 kb. This primary

library is normalized (non-normalized primary library is

NIH MGC 237) and was constructed by Express Genomics

(Frederick, MD)"

ORIGIN

Alignment Scores:

Pred. No.: 1.14e-58 Length: 777

Score: 792.50 Matches: 156

Percent Similarity: 82.0% Conservative: 31

Best Local Similarity: 68.4% Mismatches: 36

Query Match: 27.8% Indels: 6

DB: 7 Gaps: 3

US-10-635-977-2 (1-541) x CK597795 (1-777)

Qy		1	MetalloSerSeriLeuLysTrpValValSerHisGlnSer---CysSerArgSerSer	19
Db		42	GTGGCTCCGCACATCTCTCAAGTGGGTAGTCATCACCAAGAATTACTGC-----AAC	92
Qy		20	ArgSerLysProArgAspGlnArgGluAlaGlySerSerAspLeuSerSerArgGln	39
Db		93	AGAGCCAAAGGCCAAGTAAGNAGAGGAGGCGCAAGAATGGTGACCCAGCCCCCAAGAA	152
Qy		40	AspAlaGluAenAlaGluAlaLysLeuArgGlyLeuProGlyGlnLeuValAspIleAla	59
Db		153	GATCTCGATAATCTCTGACCTCAAAGTTACCAGACGCTCTCAGGACAGGTTGTAGACACAGCA	212
Qy		60	CysLysValCysGlnAlaTyrlEuGlyGlnLeuGluHisGluAspIleAspThrSerAla	79
Db		213	TGC AAGGTGTGCCAGGCCCTACTTGGACAGCTGGAGCATGAGGA CATGCACCTGTGCAGAC	272
Qy		80	AspAlaValGluAspLeuThrGluAlaGluTrpGluAspLeuThrGlnGlnTyrTyrsr	99
Db		273	CGCAGCGCCGAGGCGCTTGTCCGAGGAGGAATGGAAACGACCTGACACAACAATACTACTTGTG	332
Qy		100	LeuValHisGlyAspAlaPheIleSerAenSerArgAenTyrPheSerGlnCysGlnAla	119
Db		333	CTGGTTCATGGAATGCTTCCATACCGAATTCGAGAAGTTACTTTGCGCAGTGCCCAAGCT	392
Qy		120	LeuLeuAenArGileThrservalasnProGlnThrAspIleAspGlyLeuArgAsnIle	139
Db		393	CTGCTGTAGTAAGATCAGCTCTGTGAACCCCAAGCCGAGATTTGACGGGATACGGAACATC	452
Qy		140	TrpIlelleLysProAlaAlaLysSerArgGlyArgAspIleValCysMetAspArgVal	159
Db		453	TGGATCATTTAAGCCCTGCAGCCCAAGTCCCGGGCCCGAGATATTTGTGTGCATGGACCGGTGTG	512
Qy		160	GluGluIleLeuGluLeuAlaAlaAspHisProLeuSerArgAspAsnLysTrpVal	179
Db		513	GAGAACATCTCTGGATCTGGTGCCACAGACGCCAGACCCACGAGGACAACAATGGGTG	572
Qy		180	ValGlnLysTyrIleGluThrProLeuLeulleCysAspThrLysPheAspIleArgGln	199
Db		573	GTCCAAGAAGTACATCGAGACCCCGATGTCTATCTATGACCAACAAGTTTGACATACACAG	632
Qy		200	TrpPheLeuValThrAspTrpAsnProLeuThrIleTrpPheTyrLysGluSerTyrLeu	219
Db		633	TGGTTCCTTGTGCACAGCTGGAAATCCCCTTAACATC-TGGTTCCTACAGAGAGGCTACCTCG	691
Qy		220	ArgPheSerThrGlnArgPheSer	227
Db		692	GGTCTCTC---ACGCAGCGCTTCTCC	712

RESULT 15					
CJ018705	CJ018705	680 bp	mRNA	linear	EST 22-OCT-2004
LOCUS	CJ018705	full-length enriched swine	CDNA library,	adult spleen	Sus
DEFINITION	CJ018705	scrofa cDNA clone SFL01F010063	5',	mRNA sequence.	

ACCESSION	CJ018705	
VERSION	CJ018705.1	GI:54519944
KEYWORDS	EST.	
SOURCE	Sus scrofa (pig)	
ORGANISM	Sus scrofa	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Suina; Suidae; Sus.	
REFERENCE	1 (bases 1 to 680)	
AUTHORS	Unishi, H., Eguchi, T., Suzuki, K., Sawazaki, T., Toki, D., Shinkai, H., Okumura, N., Hamasima, N. and Awata, T.	
TITLE	PEDE (Pig EST Data Explorer): construction of a database for ESTs derived from porcine full-length cDNA libraries	
JOURNAL	Nucleic Acids Res. 32 (1), D484-D488 (2004)	
PubMed	14681463	
COMMENT	Contact: Hirohide Unishi Animal Genome Laboratory, Genome Research Department	

National Institute of Agrobiological Sciences
2 Ikenodai, Tsukuba, Ibaraki 305-8602, Japan
Tel: +81-29-838-8627
Fax: +81-29-838-8627
Email: huenishi@affrc.go.jp
EST project with full-length enriched cDNA libraries carried out in
Animal Genome Research Program (Japan) by National Institute of
Agrobiological Sciences and STAFF-Institute
Single pass sequencing of clones derived from oligo-capped cDNA
library
Vector sequences were eliminated by RepeatMasker version 2002/07/13
and crossmatch version 0.990319
Low quality bases were trimmed based on the quality values.

FEATURES	source	Location/Qualifiers
1..680	/organism="Sus scrofa"	
	/mol_type="mRNA"	
	/db_xref="taxon:9823"	
	/clone="SPL01F010063"	
	/tissue_type="spleen"	
	/dev_stage="adult"	
	/clone_lib="full-length enriched swine cDNA library, adult spleen"	
ORIGIN		
Alignment Scores:	Length: 680	
Pred. No.:	1.06e-56	
Score:	769.00	136
Percent Similarity:	78.7%	Conservative: 41
Best Local Similarity:	60.4%	Mismatches: 44
Query Match:	26.9%	Indels: 4
DB:	7	Gaps: 1
US-10-635-977-2 (1-541) x CU018705 (1-680)		
Qy	155	CysMetAspArgValGluGluLeuAlaAlaAlaAspHisProLeuSerArg 174
Db	3	TGCATGCACCACTGGAGGAGATGCTGAAGCTGGTGACAGCAACCCCATGATGATGAAG 62
Qy	175	AspAsnLysTrpValValGlnLysTyrIleGluThrProLeuLeuIleCysAspThrIys 194
Db	63	GACCGTAAATGGGTGGTGGCAAGATATATTGAGCGGCCCTCTGCTCATCTTTGGCACCAGA 122
Qy	195	PheAspIleArgGlnTrpPheLeuValThrAspTrpAsnProLeuThrIleTyrPheTyr 214
Db	123	TTTGATCTGAGACAGTGGTTCCTGGTGACCGGACTGGAAACCCACTTACTGTGCTGGTCTTAC 182
Qy	215	LysGluSerTyrLeuArgPheSerThrGlnArgPheSerLeuAspLysLeuAspSerAla 234
Db	183	CGCGACAGCTACATCCGCTTCTCCACACAGCCCTTCTCCCTGGAAGAACCTTGGACAACCTCT 242
Qy	235	IleHisLeuCysAsnAsnAlaValGlnLysTyrLeuLysAsnAspValGlyArgSerPro 254
Db	243	GTGCACCTGTGCACAACTCCATCCAGACACCTGGAGAACTCGTGCCACCGCACCCG 302
Qy	255	LeuLeuProAlaHisAsnMetTrpThrSerThrArgPheGlnGluTyrLeuGlnArgGln 274
Db	303	CTGTGGCTCGGACACACATGTGTGCGAAGAAGTTCCAAAGCCCACTTGCAGAGATT 362
Qy	275	GlyArgGlyAlaValTrpGlySerValIleTyrProSerMetLysLysAlaIleAlaHis 294
Db	363	GGGGCCCCAAACGGCTGGTCAACTATCATCTGTGCTGGCATGAAGCGCGGTGATCCAT 422
Qy	295	AlaMetLysValAlaGlnAspHisValGluProArgLysAsnSerPheGluLeuTyrGly 314
Db	423	GCCCTCCAGCGTCCGAGGACACCGTGCATGTGCGAAGCCAGCTTCCAACCTCTACGCGC 482
Qy	315	AlaAspPheValLeuGlyArgAspPheArgProTrpLeuIleGluIleAsnSerSerPro 334
Db	483	GCTGACTTTGTTATTCGGGTGAGGACTTCCAGCCCTGGCTGATCGAGATCAATGCCAGCCCC 542
Qy	335	ThrMetHisProSerThrProValThrAlaGlnLeuCysAlaGlnValGlnGluAspThr 354


```

Db      543  ACCATGGCGCCCTCCACGGCGGTCACTGCCCCGGCTCTGTGCTGGAGTGCAGCTGACACA 602
Qy      355  IleLysValAlaVal-----AspArgSerCysAspIleGlyAsnPheGluLeu 370
Db      603  CTGCGTGTGTCTCATCGACCGCGGACTGGACCGCAACTGTGTACACGGGAGCCTTTGAGCTC 662
Qy      371  LeuTrpArgGlnPro 375
Db      663  ATCTACAAGCAGCCT 677

```

Search completed: April 4, 2006, 11:34:23
Job time : 6008 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 4, 2006, 09:36:06 ; Search time 779 Seconds
(without alignments)
4628.496 Million cell updates/sec

Title: US-10-635-977-2

Perfect score: 2854

Sequence: 1 MASSILKVVVSHQSCSRSSR.....LRGLTKAEGALRPPPGKGKS 541

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-Q=/abs/ABSSWEB spool/US10635977/runat_04042006_103529_9972/app query.fasta.1
-DB=N Geneseq -QFMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=biom62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000 -HOST=abss03p
-USER=US10635977 @CGN_1_1.1096 @runat_04042006_103529_9972 -NCPU=6 -ICPU=3
-NO MAP -NEG_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N Geneseq 21.*
1: Geneseqn1980s.*
2: Geneseqn1990s.*
3: Geneseqn2000s.*
4: Geneseqn2001as.*
5: Geneseqn2001bs.*
6: Geneseqn2002as.*
7: Geneseqn2002bs.*
8: Geneseqn2003as.*
9: Geneseqn2003bs.*
10: Geneseqn2003cs.*
11: Geneseqn2003ds.*
12: Geneseqn2004as.*
13: Geneseqn2004bs.*
14: Geneseqn2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2854	100.0	1838	12 ADJ93357	Adj93357 Human BGS
2	2843	99.6	3554	12 ADJ93364	Adj93364 Human tub
3	2698.5	94.6	1859	12 ADJ93362	Adj93362 Human BGS
4	2693	94.4	3465	12 ADJ93363	Adj93363 Human BGS

c	5	2689	94.2	1939	12	ADJ93361	Adj93361 Human BGS
	6	1802	63.1	101270	12	ADQ17814	Adq17814 Human sof
	7	1299	45.5	726	12	ADJ93359	Adj93359 Human BGS
	8	1070.5	37.5	2380	6	AAS99894	Aas99894 Polynucle
	9	1054	36.9	2326	4	AAH16735	Aah16735 Human cdn
	10	1020	35.7	2538	7	ADS73168	AdS73168 Human kid
	11	1020	35.7	2538	7	ADW42022	Adw42022 cDNA elev
	12	1020	35.7	2553	13	ACN37881	Acn37881 Tumour-as
	13	1020	35.7	2881	3	AACT7214	Aac7214 Human ORF
	14	1020	35.7	3001	4	AAI58606	Aai58606 Human pol
	15	1020	35.7	3001	5	ADO98824	Ado98824 DNA encod
	16	1020	35.7	3001	9	ADB48584	AdB48584 Novel hum
	17	1012	35.5	3294	8	ABX34502	Abx34502 Human mdd
	18	998	35.0	2611	3	AAF21812	Aaf21812 Human bre
	19	983.5	34.5	2848	11	ADM03081	Adm03081 Human cdn
	20	940.5	33.0	5282	12	ADQ64426	Adq64426 Novel hum
	21	911.5	31.9	2412	12	ADH45468	Adh45468 Human mol
	22	897	31.4	2469	12	ADH45478	Adh45478 Human mol
	23	781.5	27.4	2217	10	ACA92443	AcA92443 DNA encod
c	24	772	27.0	3828	13	ADR07582	Adr07582 Full leng
c	25	760	26.6	490	12	ACH91699	Ach91699 Human gen
	26	733	25.7	2979	4	ABL17657	AbL17657 Drosophil
c	27	556.5	19.5	5728	4	ABL17656	AbL17656 Drosophil
	28	536.5	18.8	2543	4	ABL17449	AbL17449 Drosophil
	29	519.5	18.2	2044	4	AAI60392	Aai60392 Human pol
	30	516	18.1	1206	13	ADU02015	AdU02015 Novel hum
	31	514.5	18.0	4615	4	ABL17448	AbL17448 Drosophil
	32	489.5	17.2	1567	8	ACC46592	Acc46592 Human dit
	33	480.5	16.8	1958	4	AAH18689	Aah18689 Human cdn
c	34	478.5	16.8	10468	4	AAK77961	Aak77961 Human imm
	35	405	14.2	418	8	ABX52103	Abx52103 Bovine ES
	36	394	13.8	1728	5	AAS70158	Aas70158 DNA encod
	37	390.5	13.7	2250	5	AAS69383	Aas69383 DNA encod
	38	381.5	13.4	1085	4	AAF63819	Aaf63819 Human sec
	39	336	11.8	1579	8	ABX05228	Abx05228 Human nov
	40	328	11.5	1571	4	ABL14515	AbL14515 Drosophil
	41	320	11.2	2259	12	ADQ64293	AdQ64293 Novel hum
	42	318	11.1	2417	10	ADC30440	AdC30440 Human nov
	43	307.5	10.8	3141	4	ABL06959	AbL06959 Drosophil
	44	305	10.7	5314	4	AAI58261	Aai58261 Human pol
	45	305	10.7	5314	5	AQD98468	Adq98468 DNA encod

ALIGNMENTS

RESULT 1

ADJ93357

ID ADJ93357 standard; cDNA; 1838 BP.

XX

AC ADJ93357;

XX

DT 06-MAY-2004 (first entry)

XX

DE Human BGS-42 cDNA sequence SeqID1.

XX

testis-specific tubulin tyrosine-ligase-like polypeptide;
BGS-42 polypeptide; cytosolic; respiratory-Gen; gastrointestinal-Gen;
neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;
osteopathic; nootropic; antiparkinsonian; antiarthritic; antisthmatic;
anti-HIV; antibacterial; immunosuppressive; antiseborrheic;
dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;
tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;
testicular disorder; testicular cancer; pulmonary disorder; lung cancer;
gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;
brain cancer; liver cancer; proliferative condition; testis; lung;
small intestine; brain; lymph tissue; infertility; Cushing's syndrome;
emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;
Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;
sepsis; acne; Sjogren's disease; scleroderma; human; gene; ss.

XX Homo sapiens.

OS

XX

FH

Location/Qualifiers

CDS 153..1778
 FT /*tag= a
 FT /product= "Human BGS-42 protein"
 XX
 XX WO2004005487-A2.
 XX 15-JAN-2004.
 XX
 XX 09-JUL-2003; 2003WO-US021605.
 XX PR 09-JUL-2002; 2002US-0394725P.
 XX (BRIM) BRISTOL-MYERS SQUIBB CO.
 XX PI Feder JN, Wu S, Nelson TC;
 XX WPI; 2004-099381/10.
 XX DR P-PSDB; ADJ93358.
 XX
 XX New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,
 PT useful for preventing, treating or ameliorating a medical condition, e.g.
 PT aberrant cellular proliferation, reproductive disorders or testicular
 PT disorders.
 XX
 XX Claim 1; SEQ ID NO 1; 343pp; English.
 XX
 XX This invention relates to a novel testis-specific tubulin tyrosine-ligase
 CC -like polypeptide, designated the BGS-42 polypeptide. The invention may
 CC be useful for the development of compounds with a cytostatic, respiratory
 CC -Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,
 CC antiinflammatory, anabolic, hypertensive, osteopathic, neurotropic,
 CC antiparkinsonian, antiarthritic, antiaesthetic, anti-HIV, antibacterial,
 CC immunosuppressive, antiseborrheic or dermatological activity acting as
 CC tyrosine ligase modulators. In addition, the disclosed sequences may be
 CC useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be
 CC used for diagnosing a pathological condition or a susceptibility to a
 CC pathological condition in a subject, and for preventing, treating or
 CC ameliorating a medical condition, such as a disorder related to aberrant
 CC tubulin ligase activity, a disorder related to aberrant tubulin-
 CC carboxypeptidase activity, aberrant cellular proliferation, reproductive
 CC disorders, testicular disorders, testicular cancer, pulmonary disorders,
 CC lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
 CC neural disorders, brain cancer, liver cancer, or proliferative condition
 CC of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
 CC polypeptide, polynucleotide, or their modulators are also useful for
 CC treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
 CC disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
 CC -42 polypeptide can be used as a preventive agent for immunological
 CC disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
 CC disease or scleroderma. The antibodies may be used to purify, detect and
 CC target the BGS-42 polypeptides. The present sequence is that of a cDNA
 CC which encodes the human BGS-42 protein of the invention.
 XX
 XX Sequence 1838 BP; 381 A; 576 C; 583 G; 298 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 1.33e-176 Length: 1838
 Score: 2854.00 Matches: 541
 Percent Similarity: 100.0% Conservative: 0
 Best Local Similarity: 100.0% Mismatches: 0
 Query Match: 100.0% Indels: 0
 DB: 12 Gaps: 0

US-10-635-977-2 (1-541) x ADJ93357 (1-1838)

Qy 1 MetAlaSerSerIleLeuLysTropValValSerHisGlnSerCysSerArgSerArg 20
 Db 153 ATGGCATCCAGACATCCCTCAAGTGGGTGAGCCACCCAGAGCTGCACGAGGAGCAGAGA 212
 Qy 21 SerLysProArgAspGlnArgGluGluAlaGlySerSerAspLeuSerSerArgGlnAsp 40
 Db 213 AGCANGCCCGAGGACGAGGGAGGAGCCGGGAGCAGCGACCTGACGAGCAGGCAAGAT 272

Qy 41 AlaGluAsnAlaGluAlaLysLeuArgGlyLeuProGlyGlnLeuValAspIleAlaCys 60
 Db 273 GCTGAAAATGCTGAGGCAAGCTCAGGGCCCTCCCGGGGACGCTTGTGACATCGCGTGC 332
 Qy 61 LysValCysGlnAlaTyrLeuGlyGlnLeuGluHisGluAspIleAspThrSerAlaAsp 80
 Db 333 AAGGTGTGCCAGCCCTACCTGGGCGAGCTGGAGCATGAGGACATCGACACGTCAGCAGAT 392
 Qy 81 AlaValGluAspLeuThrGluAlaGluTyrGluAspLeuThrGlnGlnTyrTyrSerLeu 100
 Db 393 GCCGTGGAGACCTCACTCAGGCCGAGTGGAGGACCTGACCCAGCAGTACTACTCCCTC 452
 Qy 101 ValHisGlyAspAlaPheIleSerAsnSerArgAsnTyrPheSerGlnCysGlnAlaLeu 120
 Db 453 GTTCATGGCGATGCTTTCATCTCCAAATTCAGAAATTACTTTTCGAGTGGCAGGCTCTG 512
 Qy 121 LeuAsnArgIleThrSerValAsnProGlnThrAspIleAspGlyLeuArgAsnIleTyr 140
 Db 513 CTGNATAGAAATCAGCTCTGTGAACCTCAGACGGACATTCGCGGCTCCGGAACATCTGG 572
 Qy 141 IleIleLysProAlaAlaLysSerArgGlyArgAspIleValCysMetAspArgValGlu 160
 Db 573 ATTATAAGCCCGCGCCCAAGTCCCGGGCCGAGACATAGTGTGCATGACCGTGTGGAG 632
 Qy 161 GluIleLeuGluLeuAlaAlaAspHisProLeuSerArgAspAsnLysTyrValVal 180
 Db 633 GAGATCCTCGAGCTGGCAGCTGCAGACCACTCTTTCCAGGGACAAAGTGGGTGTC 692
 Qy 181 GlnLysTyrIleGluThrProLeuLeuIleCysAspThrLysPheAspIleArgGlnTyr 200
 Db 693 CAGAGTACATCAGACGCGCTGCTCATCTGTGCACCAAGTTCGACATCAGACAGTGG 752
 Qy 201 PheLeuValThrAspTyrAsnProLeuThrIleTyrPheTyrLysGluSerTyrLeuArg 220
 Db 753 TTCCTCGTCAGCAGCTGGAAACCCCTGACCATCTGGTTCTTACAGAGAGATTACTTGC 812
 Qy 221 PheSerThrGlnArgPheSerLeuAspLysLeuAspSerAlaIleHisLeuCysAsnAsn 240
 Db 813 TTCCTCAACTCAGCGCTTCCTCCCTGGACAAGCTGGACAGCGCCATCCACCTGTGCAAC 872
 Qy 241 AlaValGlnLysTyrLeuLysAsnAspValGlyArgSerProLeuLeuProAlaHisAsn 260
 Db 873 GCGCTCCAGAGTACCTGAAGAATGATGTGGCGCGCAGCCCTGCTGCGCGCACACAAC 932
 Qy 261 MetTyrThrSerThrArgPheGlnGluTyrLeuGlnArgGlnGlyArgGlyAlaValTyr 280
 Db 933 ATGTGACCCAGCAGCACCAGGTTCAGAGGATACCTGCAGCGCCAGGCGCGCGGTGTGG 992
 Qy 281 GlySerValIleTyrProSerMetLysLysAlaIleAlaHisAlaMetLysValAlaGln 300
 Db 993 GGCAGCGTATCTACCCCTGTCATGAAGAGGCCATCGCCACGCCATGAAGTGGCCCGAG 1052
 Qy 301 AspHisValGluProArgLysAsnSerPheGluLeuTyrGlyAlaAspPheValLeuGly 320
 Db 1053 GACCACGTGGAGCTCGAAGAACAGCTTTGAGCTCTACGCGGCTGACTTCGCTCTGGG 1112
 Qy 321 ArgAspPheArgProTyrLeuIleGluLeuAsnSerSerProThrMetHisProSerThr 340
 Db 1113 AGGGACTTCAGGCGCTCGGCTGATCGAGATCAATTCAGCGCCCACTGCACCGCTCCAGC 1172
 Qy 341 ProValThrAlaGlnLeuCysAlaGlnValGlnGluAspThrIleLysValAlaValAsp 360
 Db 1173 CCGGTACGCGCCCGAGCTGTGTGCAGAGTGCAGAGGACACCATCAAGTGGCGGTGGAC 1232
 Qy 361 ArgSerCysAspIleGlyAsnPheGluLeuTyrArgGlnProValValGluProPro 380
 Db 1233 CGCAGCTGTGACATCGCAACTTCGAGCTCTGTGGAGGAGCGCGGTGTGAGCCGCC 1292
 Qy 381 ProPheSerGlySerAspLeuCysValAlaGlyValSerValArgAlaAArgArgGln 400
 Db 1293 CCATTACGCGGGTCCGACCTCTCGTGGCGGGCGTCAAGTGTGAGGAGAGCCAGAGGCGAG 1352
 Qy 401 ValLeuProValCysAsnLeuLysAlaSerAlaSerLeuLeuAspAlaGlnProLeuLys 420

DB 1353 GTGCTCCGCTGTGCAACCTCAAGGCTCGCCCTCGCTGTGGACCGCGCTGAAG 1412
QY 421 AlaArgGlyProSerAlaMetProAspProAlaGlnGlyProProSerProAlaLeuGln 440
DB 1413 GCACGGGGCCCTCGGCCATGCGACACCTGCGCCAGGGACCCCATCACAGCTCTCCAG 1472
QY 441 ArgAspLeuGlyLeuGlyGluGlyGlyGlyLeuProLeuAlaLeuAlaProLeuArg 460
DB 1473 CGGGACTTGGGACTGAAGAAGAGAGGGGCTCCCTCGCTGTGGCAGCCCTTAAGG 1532
QY 461 GlyAlaAlaGluSerGlyGlyAlaAlaGlnProThrArgThrLysAlaAlaGlyLysVal 480
DB 1533 GGGGACGCGAGGCGGTGGAGCGGCACAGCCACCCGACCCAACTGCTGGGAAGGTG 1592
QY 481 GluLeuProAlaCysProCysArgHisValAspSerGlnAlaProAsnThrGlyValPro 500
DB 1593 GAGTCCCGGCTCGCCCTGTGCGCAGTGGACAGTCAGGCCCAACACCGGTGTCCTC 1652
QY 501 ValAlaGlnProAlaLysSerTrpAspProAsnGlnLeuAsnAlaHisProLeuGluPro 520
DB 1653 GTAGCCACCGCCGCAAAAGCTGGATCCAAACCCAGCTAAATCGCACCGCTGGAGCT 1712
QY 521 ValLeuArgGlyLeuLysThrAlaGluGlyAlaLeuArgProProGlyGlyLysGly 540
DB 1713 GTGTGCGGGCTGGAAGACAGAGGGCGCGCTCGCTCGCGCCCGCGAGAAAGT 1772
QY 541 Ser 541
DB 1773 TCA 1775

RESULT 2

ID ADJ93364 standard; DNA; 3554 BP.

AC ADJ93364;

DT 06-MAY-2004 (first entry)

DE Human tubulin tyrosine ligase protein consensus gene sequence SeqID12.

KW testis-specific tubulin tyrosine-ligase-like polypeptide;
KW BGS-42 polypeptide; cytostatic; respiratory-Gen; gastrointestinal-Gen;
KW neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;
KW osteopathic; nootropic; antiparkinsonian; antiarthritic; antiasthmatic;
KW anti-HIV; antibacterial; immunosuppressive; antiseborrheic;
KW dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;
KW tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;
KW testicular disorder; testicular cancer; pulmonary disorder; lung cancer;
KW gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;
KW brain cancer; liver cancer; proliferative condition; testis; lung;
KW small intestine; brain; lymph tissue; infertility; Cushing's syndrome;
KW emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;
KW Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;
KW sepsis; acne; Sjogren's disease; scleroderma; human; gene; ds.
XX Homo sapiens.
OS Synthetic.

XX Key Location/Qualifiers

FT 1695..3320

FT /*tag= a

FT /product= "Human tubulin tyrosine ligase protein"

XX W02004005487-A2.

XX 15-JAN-2004.

XX 09-JUL-2003; 2003WO-US021605.

XX 09-JUL-2002; 2002US-0394725P.

XX (BRIM) BRISTOL-MYERS SQUIBB CO.

XX Feder JN, Wu S, Nelson TC;
XX WPI; 2004-099381/10.
DR P-PSDB; ADJ93365.
XX New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,
PT useful for preventing, treating or ameliorating a medical condition, e.g.
PT aberrant cellular proliferation, reproductive disorders or testicular
PT disorders.
XX Example 4; SEQ ID NO 12; 343pp; English.
XX This invention relates to a novel testis-specific tubulin tyrosine-ligase
CC -like polypeptide, designated the BGS-42 polypeptide. The invention may
CC be useful for the development of compounds with a cytostatic, respiratory
CC -Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,
CC antiinflammatory, anabolic, hypertensive, osteopathic, nootropic,
CC antiparkinsonian, antiarthritic, antiasthmatic, anti-HIV, antibacterial,
CC immunosuppressive, antiseborrheic or dermatological activity acting as
CC tyrosine ligase modulators. In addition, the disclosed sequences may be
CC useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be
CC used for diagnosing a pathological condition or a susceptibility to a
CC pathological condition in a subject, and for preventing, treating or
CC ameliorating a medical condition, such as a disorder related to aberrant
CC tubulin ligase activity, a disorder related to aberrant tubulin-
CC carboxypeptidase activity, aberrant cellular proliferation, reproductive
CC disorders, testicular disorders, testicular cancer, pulmonary disorders,
CC lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
CC neural disorders, brain cancer, liver cancer, or proliferative condition
CC of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
CC polypeptide, polynucleotide, or their modulators are also useful for
CC treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
CC disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
CC -42 polypeptide can be used as a preventive agent for immunological
CC disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
CC disease or scleroderma. The antibodies may be used to purify, detect and
CC target the BGS-42 polypeptides. The present sequence is that of the
CC tubulin tyrosine ligase protein consensus gene sequence which was used in
CC the exemplification of the invention.
XX SQ Sequence 3554 BP; 692 A; 1090 C; 1157 G; 615 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 1.49e-175 Length: 3554
Score: 2843.00 Matches: 539
Percent Similarity: 99.6% Conservatives: 0
Best Local Similarity: 99.6% Mismatches: 2
Query Match: 99.6% Indels: 0
DB: 12 Gaps: 0

US-10-635-977-2 (1-541) x ADJ93364 (1-3554)

QY 1 MetAlaSerSerIleLeuLysTrpValValSerHisGlnSerCysSerArgSerArg 20
DB 1695 ATGGCATCCAGCATCTCTCAAGTGGTGTGTGACCCAGAGCTGCAGGAGCAGCAGA 1754
QY 21 SerLysProArgAspGlnArgGluAlaGlySerSerAspLeuSerSerArgGlnAsp 40
DB 1755 AGCAAGCCAGGAGCAGGAGGAGGCGGAGCAGCAGCTGAGCAGGAGGCAAGAT 1814
QY 41 AlaGluAsnAlaGluAlaLysLeuArgGlyLeuProGlyGlnLeuValAspIleAlaCys 60
DB 1815 GCTGAATGCTGAGGCAAAAGCTCAGGGGCGCTCCCGGGGAGCTTGTGACATCGGTGC 1874
QY 61 LysValCysGlnAlaTyrLeuGlyGlnLeuGluHisGluAspIleAspThrSerAlaAsp 80
DB 1875 AAGGTGTCAGGCGCTACCTGGGGCAGCTGGAGCATGAGGACATGACACGCTCAGCAGAT 1934
QY 81 AlaValGluAspLeuThrGluAlaGluTrpGluAspLeuThrGlnGluTyrSerLeu 100
DB 1935 GCCGTGGAGGAGCTCCTGAGGCGGAGTGGGAGGAGCTGACCCAGCAGTACTACTCCCTC 1994

Qy	101	ValHisGlyAspAlaPheIleSerAsnSerArgAsnTyrPheSerGlnCysGlnAlaLeu	120
Db	1995	GTTCATGGCGATGCTTTTCATCTCCAAATTCAGAAATTTACTTTTCGAGTCGCCAGGCTCTG	2054
Qy	121	LeuAsnArgIleThrSerValAsnProGlnThrAspIleAspGlyLeuArgAsnIleTrp	140
Db	2055	CTGAATAGAAATCAGCTCTGTAACCCCTCAGACGGACATTACCGGGCTCCGGAAACATCTGG	2114
Qy	141	IleIleLysProAlaAlaLysSerArgGlyArgAspIleValCysMetAspArgValGlu	160
Db	2115	ATTATAAAGCCGGCGCCAGTCCCGGGGCGAGACATAGTGTGCATGGACCGTGTGGAG	2174
Qy	161	GluIleLeuGluLeuAlaAlaAspHisProLeuSerArgAspAsnLysTrpValVal	180
Db	2175	GAGATCTGGAGCTGGCAGCTGCAGACCAACCTCTTTCCAGGGACAAACAAGTGGGTGTC	2234
Qy	181	GlnLysTrpIleGluThrProLeuLeuIleCysAspThrLysPheAspIleArgGlnTrp	200
Db	2235	CAGAAGTACATCGAGACCGCGTCTCATCTGTGTACACCAAGTTTCGACATCAGACAGTGG	2294
Qy	201	PheLeuValThrAspTrpAsnProLeuThrIleTrpPheTyrLysGluSerTyrLeuArg	220
Db	2295	TTCTCTGTCAGGACTGGAAACCCCTGACCATCTGGTTCTACAGGAGAGTTACTTCGCG	2354
Qy	221	PheSerThrGlnArgPheSerLeuAspLysLeuAspSerAlaIleHisLeuCysAsnAsn	240
Db	2355	TTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGACAGCGCCATCCACCTGTGCAACAAC	2414
Qy	241	AlaValGlnLysTyrLeuLysAsnAspValGlyArgSerProLeuLeuProAlaHisAsn	260
Db	2415	GCCGTCAGCAAGTAGCTGAAGATGATGTGGCGCGACGCCCTCTGTCGCCGACACACAAC	2474
Qy	261	MetTrpThrSerThrArgPheGlnGluTyrLeuGlnArgGlnGlyArgGlyAlaValTrp	280
Db	2475	ATGTGGACACAGACACAGGTTCCAGAGTACTCTGCAGCGCACAGGGCCCTGGCGCGTGG	2534
Qy	281	GlySerValIleTyrProSerMetLysLysAlaIleAlaHisAlaMetLysValAlaGln	300
Db	2535	GGCAGGTCATCTACCCGTCATGAAGAGGCCATCGCCACGCCCATGAAGTGGCCCGAC	2594
Qy	301	AspHisValGluProArgLysAsnSerPheGluLeuTyrGlyAlaAspPheValLeuGly	320
Db	2595	GACCACGTGGAGCCTCGCAAGAACAGCTTTAGAGTCTACGGGGGTGACTTCGTCTTTGGG	2654
Qy	321	ArgAspPheArgProTrpLeuIleGluIleAsnSerSerProThrMetHisProSerThr	340
Db	2655	AGGGACTTCAGGCCCTGGCTGATCGAGATCAATTTCCAGCCCCACATGCACCCGCTCCAG	2714
Qy	341	ProValThrAlaGlnLeuCysAlaGlnValGlnGluAspThrIleLysValAlaValAsp	360
Db	2715	CCGGTCAGGGCCAGCTGTGTGCACAGGTGCAGGAGGACCATCAAGTGGCGCTGGAC	2774
Qy	361	ArgSerCysAspIleGlyAsnPheGluLeuLeuTrpArgGlnProValValGluProPro	380
Db	2775	CGCAGCTGTGACATCGGCCAATTCTGAGTGGAGGCCAGCCGCTGGTTGAGCCGCC	2834
Qy	381	ProPheSerGlySerAspLeuCysValAlaGlyValSerValArgAlaArgArgGln	400
Db	2835	CCATTACGGGGTCGACCTCTGGGTGGCGGGCTGAGTGTGAGAGAGCCAGAGGGCAG	2894
Qy	401	ValLeuProValCysAsnLeuLysAlaSerAlaSerLeuLeuAspAlaGlnProLeuLys	420
Db	2895	GTGCTGCCCGTCTGCNACTTCAGGCCTCGGCCCTCGCTGTGGACGCGCAGCCGCTGAAG	2954
Qy	421	AlaArgGlyProSerAlaMetProAspProAlaGlnGlyProProSerProAlaLeuGln	440
Db	2955	GCACGGGGCCCTCGGCCATGCCAGACCTCGCCAGGGACCCCCATCAACAGCTCTCCAG	3014
Qy	441	ArgAspLeuGlyLeuLysGluLysGlyValLeuProLeuAlaLeuLeuAlaProLeuArg	460
Db	3015	CGGGACTTTGGGACTGAAGGAGAGAAAGGGGCTCCCCCTGGCGCTTCTCTGGCACCTTAAG	3074
Qy	461	GlyAlaAlaGluSerGlyGlyValAlaAlaGlnProThrArgThrLysAlaAlaGlyLysVal	480

Db	3075	GGGGCAGCCGAGAGCGGTGGAGCGGCACAGCCACCCGACCAAGAGCTCTGGAAAGTG	3134
Qy	481	GluLeuProAlaCysProCysArgHisValAspSerGlnAlaProAsnThrGlyValPro	500
Db	3135	GAGCTCCGGCCCTGCCCTGTCCGACAGTGGACAGTCAGGCCCCAAACACCGGTGTCCCC	3194
Qy	501	ValAlaGlnProAlaIysSerTrpAspProAsnGlnIleuAsnAlaHisProLeuGluPro	520
Db	3195	GTAGCCAGCCGCCCAAAAGCTGGGATCCAAACACAGCTAAATGAGCACCCGCTGGAGCCT	3254
Qy	521	ValIleuArgGlyIleuIysThrAlaGluGlyAlaLeuArgProProP-oGlyGlyIysGly	540
Db	3255	GTCTCGGAGCCTGAAGACAGCAGAGGGCGCGCTGCTCGCGCCCGGAGAAAGGT	3314
Qy	541	Ser	541
Db	3315	TCA	3317
RESULT 3			
ADJ93362			
ID	ADJ93362	standard; DNA; 1859 BP.	
XX	AC	ADJ93362;	
AC			
XX			
DT	06-MAY-2004	(first entry)	
XX			
DE	Human BGS-42	protein-related DNA clone B SeqID10.	
XX			
KW	testis-specific tubulin tyrosine-ligase-like polypeptide;		
KW	BGS-42 polypeptide; cytosolic; respiratory-Gen; gastrointestinal-Gen;		
KW	neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;		
KW	osteopathic; nontropic; antiparkinsonian; antiarthritic; antiasthmatic;		
KW	anti-HIV; antibacterial; immunosuppressive; antiiseborrheic;		
KW	dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;		
KW	tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;		
KW	testicular disorder; testicular cancer; pulmonary disorder; lung cancer;		
KW	gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;		
KW	brain cancer; liver cancer; proliferative condition; testis; lung;		
KW	small intestine; brain; lymph tissue; infertility; Cushing's syndrome;		
KW	emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;		
KW	Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;		
KW	sepsis; acne; Sjogren's disease; scleroderma; human; ds.		
XX			
OS	Homo sapiens.		
XX			
XX	W02004005487-A2.		
PN			
XX			
PD	15-JAN-2004.		
XX			
PF	09-JUL-2003; 2003WO-US021605.		
XX			
XX	09-JUL-2002; 2002US-0394725P.		
FR			
XX	(BRIM) BRISTOL-MYERS SQUIBB CO.		
FA			
XX			
PI	Feder JN, Wu S, Nelson TC;		
XX			
XX	WPI; 2004-099381/10.		
DR			
XX			
PT	New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,		
PT	useful for preventing, treating or ameliorating a medical condition, e.g.		
PT	aberrant cellular proliferation, reproductive disorders or testicular		
PT	disorders.		
XX			
PS	Example 4; SEQ ID NO 10; 343pp; English.		
XX			
CC	This invention relates to a novel testis-specific tubulin tyrosine-ligase		
CC	-like polypeptide, designated the BGS-42 polypeptide. The invention may		
CC	be useful for the development of compounds with a cytosolic, respiratory		
CC	-Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,		
CC	antiinflammatory, anabolic, hypertensive, osteopathic, nontropic,		
CC	antiparkinsonian, antiarthritic, antiasthmatic, anti-HIV, antibacterial,		

CC immunosuppressive, antiseborrheic or dermatological activity acting as
 CC tyrosine ligase modulators. In addition, the disclosed sequences may be
 CC useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be
 CC used for diagnosing a pathological condition or a susceptibility to a
 CC pathological condition in a subject, and for preventing, treating or
 CC ameliorating a medical condition, such as a disorder related to aberrant
 CC tubulin ligase activity, a disorder related to aberrant tubulin-
 CC carboxypeptidase activity, aberrant cellular proliferation, reproductive
 CC disorders, testicular disorders, testicular cancer, pulmonary disorders,
 CC lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
 CC neural disorders, brain cancer, liver cancer, or proliferative condition
 CC of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
 CC polypeptide, polynucleotide, or their modulators are also useful for
 CC treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
 CC disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
 CC -42 polypeptide can be used as a preventive agent for immunological
 CC disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
 CC disease or scleroderma. The antibodies may be used to purify, detect and
 CC target the BGS-42 polypeptides. The present sequence is that of a DNA
 CC clone sequence which is related to the invention.

XX
 SQ Sequence 1859 BP; 418 A; 563 C; 562 G; 316 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 1,81e-166 Length: 1859
 Score: 2698.50 Matches: 522
 Percent Similarity: 96.5% Conservative: 0
 Best Local Similarity: 96.5% Mismatches: 4
 Query Match: 94.6% Indels: 17
 DB: 12 Gaps: 2

US-10-635-977-2 (1-541) x ADJ93362 (1-1859)

CC	1	MetAlaSerSerIleLeuLysTrpValValSerHisGlnSerCysSerArgSerArg	20
QY			
DB	47	ATGGCATCAGCATCTCAAGTGGTGGTTCAGCCACAGAGCTGCAGCAGGAGCAGCAGA	106
QY	21	SerLysProArgAspGlnArgGluGluAlaGlySerSerAspLeuSerArgGlnAsp	40
DB	107	AGCAAGCCAGGAGCAGAGGAGGAGGCGCGAGCAGCGACCTGAGCAGCAGGCAAGAT	166
QY	41	AlaGluAsnAlaGluAlaLysLeuArgGlyLeuProGlyGlnLeuValAspIleAlaCys	60
DB	167	GCTGAATAATGTCTGAGCAAAAGCTCAGGGGCTCCCGGGGAGCTTGTGACATCGCGTGC	226
QY	61	LysValCysGlnAlaTyrLeuGlyGlnLeuGluHisGluAspIleAspThrSerAlaAsp	80
DB	227	AAGGTGTCCAGCCCTACCTGGGGCAGCTGGAGCATGAGGACATCGACACGTCAGCAGAT	286
QY	81	AlaValGluAspLeuThrGluAlaGluTrpGluAspLeuThrGlnGlnTyrTyrSerLeu	100
DB	287	GCGGTGGAGGACCTCACTGAGCCGAGTGGAGGACCTGACCCAGCAGTACTACTCCCTC	346
QY	101	ValHisGlyAspAlaPheIleSerAsnSerArgAsnTyrPheSerGlnCysGlnAlaLeu	120
DB	347	GTTC-----TGCCAGGCTCTG	363
QY	121	LeuAsnArgIleThrSerValAsnProGlnThrAspIleAspGlyLeuArgAsnIleTrp	140
DB	364	CTGAATAGAAATCAGCTGTGTGAACCTTCAGACGACATTCAGGGCTCCGGAACTCTGG	423
QY	141	IleIleLysProAlaAlaLysSerArgGlyArgAspIleValCysMetAspArgValGlu	160
DB	424	ATTATAAGACCCCGCGCCCAAGTCCCGGGCCGAGACATAGTGTGATGACCGTGTGGAG	483
QY	161	GluIleLeuGluLeuAlaAlaAspHisProLeuSerArgAspAsnLysTrpValVal	180
DB	484	CAGATCCTGGAGCTGGCAGCTGCAGACCACTCTTTCCAGGACAAACAAGTGGGTGTC	543
QY	181	GlnLysTyrIleGluThrProLeuLeuIleCysAspThrLysPheAspIleArgGlnTrp	200
DB	544	CAGAAAGTACATCGACAGCCCTGCTCATCTGTGACACCAAGTTCGACATCAGACAGTGG	603
QY	201	PheLeuValThrAspTrpAsnProLeuThrIleTrpPheTyrLysGluSerTyrLeuArg	220
DB	604	TTCTTCGTCAACGACTGGAAACCCCTTGACATCTGTTCTACAAGGAGAGTACTTTGCGG	663
QY	221	PheSerThrGlnArgPheSerLeuAspLysLeuAspSerAlaIleHisLeuCysAsnAsn	240
DB	664	TTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGACAGGCCATCCACCTGTGCAACAAC	723
QY	241	AlaValGlnLysTyrLeuLysAsnAspValGlyArgSerProLeuLeuProAlaHisAsn	260
DB	724	GCCGTCCAGAGTACTTGAAGATGATGTGGCGGAGCCCTGCTGCCGACACACAC	783
QY	261	MetTrpThrSerArgPheGlnGlnTyrLeuGlnArgGlnGlyArgGlyAlaValTrp	280
DB	784	ATGTGACACAGCAGGTTCCAGGAGTACCTGCAGCGCAGCGCGCTGGCGCGTGTGG	843
QY	281	GlySerValIleTyrProSerMetLysLysAlaIleAlaHisAlaMetLysValAlaGln	300
DB	844	GGCAGGCTCATCTACCCCTCATGAAGAGGCCATCGCCACGCCATGAAGTGGGCCAG	903
QY	301	AspHisValGluProArgLysAsnSerPheGluLeuTyrGlyAlaAspPheValLeuGly	320
DB	904	GACCATGGAGCCTCGCAAGAACAGCTTTGAGCTCTACGGGGCTGACTTGTCTCTGGG	963
QY	321	ArgAspPheArgProTrpLeuLeuGluLeuAsnSerSerProThrMetHisProSerThr	340
DB	964	AGGGACTTCAGGCCCTGGCTGATCGAGATCAATTCCAGCCCCCACCACCGTCCACG	1023
QY	341	ProValThrAlaGlnLeuCysAlaGlnValGlnGluAspThrIleLysValAlaValAsp	360
DB	1024	CCGGTCACGGCCAGCTGTGTGCACAGGTGAGGAGGACCATCAAGTGGCGGTGGAC	1083
QY	361	ArgSerCysAspIleGlyAsnPheGluLeuLeuTrpArgGlnProValValGluProPro	380
DB	1084	CGCAGCTGTGACATCGGCAACTTCGAGCTCTCTGTGAG-----CCGGTGTGTGAGCCGCC	1139
QY	381	ProPheSerGlySerAspLeuCysValAlaGlyValSerValArgAlaArgArgGln	400
DB	1140	CCATTCAGCGGTCCGACCTCTCGCTGGCGGCGCTCAGTGTGAGGAGAGCCAGGAGCAG	1199
QY	401	ValLeuProValCysAsnLeuLysAlaSerAlaSerLeuLeuAspAlaGlnProLeuLys	420
DB	1200	GTGCTGCCCTGTGCAACCTCAAGGCTCGGCTCGCTGTGTGGACGCGCAGCGCTGAAG	1259
QY	421	AlaArgGlyProSerAlaMetProAspProAlaGlnGlyProProSerProAlaLeuGln	440
DB	1260	GCACGGGGCCCTCGGCCATGCCAGACCTCGCCAGGGAGACCCCATCACAGCTCTCCAG	1319
QY	441	ArgAspLeuGlyLysGluLysGlyLeuProLeuAlaLeuAlaProLeuArg	460
DB	1320	CGGGACTTGGGACTGAGGAGAGAGAGGGGCTCCCTCCCTGGCTTGTGGACCCCTTAAGG	1379
QY	461	GlyAlaAlaGluSerGlyGlyAlaAlaGlnProThrArgThrLysAlaAlaGlyLysVal	480
DB	1380	GGGGCAGCGAGCGGTGGAGCCGACAGCCCGCACCCCAAGCTGCTGGGAAGGTG	1439
QY	481	GluLeuProAlaCysProCysArgHisValAspSerGlnAlaProAsnThrGlyValPro	500
DB	1440	GAGCTCCCGGCTCGCCCTGTGCGCACGTGGACAGTGGGGCCCAACACCGGCTGCC	1499
QY	501	ValAlaGlnProAlaLysSerTrpAspProAsnGlnLeuAsnAlaHisProLeuGluPro	520
DB	1500	GTAGCCAGCCCGCCAAAGCTGGGATCCAAACACAGCTAAATGAGCACCCCGTGGAGCT	1559
QY	521	ValLeuArgGlyLeuLysThrAlaGluGlyAlaLeuArgProProGlyGlyLysGly	540
DB	1560	GTGCTGCGGAGCTGAAGACAGCAGGCGCGCTGCTGCTCGCGCCCGCGAGGAAAAGT	1619
QY	541	Ser	541
DB	1620	TCA	1622

RESULT 4

ADJ93363
ID ADJ93363 standard; DNA; 3465 BP.
XX
AC ADJ93363;
XX
DT 06-MAY-2004 (first entry)
XX
DE Human BGS-42 protein-related DNA clone C SeqID11.
XX
KW testis-specific tubulin tyrosine-ligase-like polypeptide;
KW BGS-42 polypeptide; cytosolic; respiratory; gastrointestinal; anabolic; hypertensive;
KW neuroprotective; endocrine; anti-inflammatory; antiarthritic; antiseborrheic;
KW osteopathic; nontropic; antiparkinsonian; antiseborrheic;
KW anti-HIV; antibacterial; immunosuppressive; antiseborrheic;
KW dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;
KW tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;
KW testicular disorder; testicular cancer; pulmonary cancer; lung cancer;
KW gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;
KW brain cancer; liver cancer; proliferative condition; testis; lung;
KW small intestine; brain; lymph tissue; infertility; Cushing's syndrome;
KW emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;
KW Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;
KW sepsis; acne; Sjogren's disease; scleroderma; human; ds.
XX
OS Homo sapiens.
XX
XX W02004005487-A2.
XX
XX 15-JAN-2004.
XX
XX 09-JUL-2003; 2003WO-US021605.
XX
XX 09-JUL-2002; 2002US-0394725P.
XX
XX (BRIM) BRISTOL-MYERS SQUIBB CO.
XX
XX Feder JN, Wu S, Nelson TC;
XX
XX WPI; 2004-099381/10.
XX
XX New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,
XX useful for preventing, treating or ameliorating a medical condition, e.g.
XX PT aberrant cellular proliferation, reproductive disorders or testicular
XX disorders.
XX
XX Example 4; SEQ ID NO 11; 343pp; English.
XX
XX This invention relates to a novel testis-specific tubulin tyrosine-ligase
XX -like polypeptide, designated the BGS-42 polypeptide. The invention may
XX be useful for the development of compounds with a cytostatic, respiratory
XX -Gen, gastrointestinal, antiarthritic, osteopathic, endocrine-Gen,
XX anti-inflammatory, anabolic, hypertensive, osteopathic, nontropic,
XX antiparkinsonian, antiseborrheic or dermatological activity acting as
XX immunosuppressive, antiseborrheic or dermatological activity acting as
XX tyrosine ligase modulators. In addition, the disclosed sequences may be
XX useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be
XX used for diagnosing a pathological condition or a susceptibility to a
XX pathological condition in a subject, and for preventing, treating or
XX ameliorating a medical condition, such as a disorder related to aberrant
XX tubulin ligase activity, a disorder related to aberrant tubulin-
XX carboxypeptidase activity, aberrant cellular proliferation, reproductive
XX disorders, testicular disorders, testicular cancer, pulmonary disorders,
XX lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
XX neural disorders, brain cancer, liver cancer, or proliferative condition
XX of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
XX polypeptide, polynucleotide, or their modulators are also useful for
XX treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
XX disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
XX -42 polypeptide can be used as a preventive agent for immunological
XX disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
XX disease or scleroderma. The antibodies may be used to purify, detect and
XX target the BGS-42 polypeptides. The present sequence is that of a DNA
XX clone sequence which is related to the invention.

XX
SQ Sequence 3465 BP; 667 A; 1074 C; 1126 G; 598 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 8.51e-166 Length: 3465
Score: 2693.00 Matches: 516
Percent Similarity: 95.4% Conservatives: 0
Best Local Similarity: 95.4% Mismatches: 3
Query Match: 94.4% Indels: 23
DB: 12 Gaps: 1
US-10-635-977-2 (1-541) x ADJ93363 (1-3465)
QY 1 MetAlaSerSerIleLeuLysTrpValValSerHisGlnSerCysSerArgSerArg 20
Db 1678 ATGGCATCCAGATCTCTCAAGTGGTGGTTCAGCCACAGAGTGCAGGAGCAGCAGA 1737
QY 21 SerLysProArgAspGlnArgGluGluAlaGlySerSerAspLeuSerArgGlnAsp 40
Db 1738 AGCAAGCCAGGAGACAGAGGAGGAGGCGGAGCAGCGACCTGAGCAGCAGGCAAGAT 1797
QY 41 AlaGluAsnAlaGluAlaLysLeuArgGlyLeuProGlyGlnLeuValAspIleAlaCys 60
Db 1798 GCTGAAAATGCTGAGCAAGCTCAGGGGCTCCCGGGGAGCTTGTGACATCGCGTGC 1857
QY 61 LysValCysGlnAlaTyrLeuGlyGlnLeuGluHisGluAspIleAspThrSerAlaAsp 80
Db 1858 AAGGTGTGCCAGCCCTACTTGGGCGAGCTGGAGCATGAGGACATCGACAGTCAGCAGAT 1917
QY 81 AlaValGluAspLeuThrGluAlaGluTrpGluAspLeuThrGlnGlnTyrTyrSerLeu 100
Db 1918 GCGGTGGAGACCTCACTGAGGCGGAGTGGAGGACCTGACCCAGCAGTACTACTCCCTC 1977
QY 101 ValHisGlyAspAlaPheIleSerAsnSerArgAsnTyrPheSerGlnCysGlnAlaLeu 120
Db 1978 GTTCATGGCGATGCTTTTCATCTCCAATTCAAGAAATTACTTTTCGAGTCCAGGCTCG 2037
QY 121 LeuAsnArgIleThrSerValAsnProGlnThrAspIleAspGlyLeuArgAsnIleTyr 140
Db 2038 CTGAATAGATACACGCTGTGAACCTTCAGACGACATTTGACGGGCTCCGGACATCTGG 2097
QY 141 IleIleLysProAlaLysSerArgGlyArgAspIleValCysMetAspArgValGlu 160
Db 2098 ATTATAAGCCCGCGCCCAAGTCCCGGGCGG----- 2129
QY 161 GluIleLeuGluLeuAlaAlaAlaAspHisProLeuSerArgAspAsnLysTrpValVal 180
Db 2130 -----AGGGACCAACAAAGTGGGTGGTC 2150
QY 181 GlnLysTyrIleGluThrProLeuLeuIleCysAspThrLysPheAspIleArgGlnTyr 200
Db 2151 CAGAAGTACATCAGAGAGCGCGTCTCATCTGTGACACCAAGTTGCACATCAGACAGTGG 2210
QY 201 PheLeuValThrAspTrpAsnProLeuThrIleTyrPheTyrLysGluSerTyrLeuArg 220
Db 2211 TTCTCGTCAAGGACGAGACCCCTGACCATCTGGTTCTACAAAGGAGAGTTACTTGGCG 2270
QY 221 PheSerThrGlnArgPheSerLeuAspLysLeuAspSerAlaIleHisLeuCysAsnAsn 240
Db 2271 TTCTCAACTCAGGCGCTTCTCCCTGGCAAGCTGGACAGCGCCATCCACCTGTGCAACAC 2330
QY 241 AlaValGlnLysTyrLeuLysAsnAspValGlyArgSerProLeuLeuProAlaHisAsn 260
Db 2331 GCGGTCAGAAAGTACCTGAAGAATGATGTGGCGCGAGCCCTCTGCTGCGGACACACAC 2390
QY 261 MetTrpThrSerThrArgPheGlnGluTyrLeuGlnArgGlnGlyArgGlyAlaValTyr 280
Db 2391 ATGTGGACCAAGCAGGTTCCAGGAGTACTTGCAGCGCCAGGCGCGTGGCGCGTGG 2450
QY 281 GlySerValIleTyrProSerMetLysLysAlaIleAlaHisAlaMetLysValAlaGln 300
Db 2451 GGCAGCGTCTATACCGTCCATGAGAGAGGCCATCGCCACCGCATGAAGGTGGCCGAG 2510

QY	41	AlaGluAsnAlaGluAlaLysLeuArgGlyLeuProGlyGlnLeuValAspIleAlaCys	60	QY	401	ValLeuProValCysAsnLeuLysAlaSerLeuLeuAspAlaGlnProLeuLys	420
Db	273	GCTGAAATGCTGAGCAAGCTCAGGGGCTCCCGGGCAGCTTGTGGACATCGCGTGC	332	Db	1286	GTGCTGCCCTGTGCAACCTCAAGGCTCGGCTCGTGTGGACGCGCAGCGCTGAAG	1345
QY	61	LysValCysGlnAlaTyrLeuGlyGlnLeuGluHisGluAspIleAspThrSerAlaAsp	80	QY	421	AlaArgGlyProSerAlaMetProAspProAlaGlnGlyProProSerProAlaLeuGln	440
Db	333	AAGGTGTCCAGGCCCTACCTGGGGCAGCTGGACATGAGGACATCGACAGTCAGCAGAT	392	Db	1346	GCACGGGGCCCTCGGCCATGCCAGACCTTGCCAGGGACCCCATCACAGCTCTCCAG	1405
QY	81	AlaValGluAspLeuThrGluAlaGluTrpGluAspLeuThrGlnGlnTyrSerLeu	100	QY	441	ArgAspLeuGlyLeuLysGlnGlyLeuProLeuAlaLeuLeuAlaProLeuArg	460
Db	393	GCCGTGGAGCACCTCACTGAGGCCGAGTGGAGGACCTGCACCCAGCAGTACTCCCTC	452	Db	1406	CGGGACTTGGGACTGAAGGAAGAGAGGGGCTCCCCCTGGCTTGTGGACCTTAAAG	1465
QY	101	ValHisGlyAspAlaPheIleSerAsnSerArgAsnTyrPheSerGlnCysGlnAlaLeu	120	QY	461	GlyAlaAlaGluSerGlyGlyAlaAlaGlnProThrArgThrLysAlaAlaGlyLysVal	480
Db	453	GTTTCATGGCGATGCTTTCATCTCAATTCAGAAATTACTTTTCGAGGTGCCAGGCTCG	512	Db	1466	GGGGCAGCCAGAGCGGTGGAGCCGACAGCCACCCGACCAAGCTGCTGGGAAGGTG	1525
QY	121	LeuAsnArgIleThrSerValAsnProGlnThrAspIleAspGlyLeuArgAsnIleTyr	140	QY	481	GluLeuProAlaCysProCysArgHisValAspSerGlnAlaProAsnThrGlyValPro	500
Db	513	CTGAATAGAAATCAGTCTGTGAACCTTCAGACGGACATTGACGGGCTCCGGAAATCTGG	572	Db	1526	GAGCTCCGGGCTGCTCCCTGTGCCACGTGGACAGTCAGGCCCAACACCGGTGCCCT	1585
QY	141	IleIleLysProAlaLysSerArgGlyArgAspIleValCysMetAspArgValGlu	160	QY	501	ValAlaGlnProAlaLysSerTrpAspProAsnGlnLeuAsnAlaHisProLeuGluPro	520
Db	573	ATTATAAGCCCGGGCCAAAGTCCCGGGCGG-----	604	Db	1586	GTAGCCAGCCCGCCAAAAGCTGGGATCCAAACCACTAAATGAGCACCGCTGGAGCCT	1645
QY	161	GluIleLeuGluLeuAlaAlaAspHisProLeuSerArgAspAsnLysTrpValVal	180	QY	521	ValLeuArgGlyLeuLysThrAlaGluGlyAlaLeuArgProProGlyGlyLysGly	540
Db	605	-----AGGGACAACAAGTGGGTGGTC	625	Db	1646	GTGCTCGGAGGCTGAAGACAGCAGAGGGCGGCTCGCTCCGCCGCCGAGGAAAAGGT	1705
QY	181	GlnLysTyrIleGluThrProLeuLeuIleCysAspThrLysPheAspIleArgGlnTrp	200	QY	541	Ser 541	
Db	626	CAGAAGTACATCAGACGCGCTGCTCATCTGTGACACCAAGTTCGACATCAGACAGTGG	685	Db	1706	TCA 1708	
QY	201	PheLeuValThrAspTrpAsnProLeuThrIleTrpPheTyrLysGluSerTyrLeuArg	220	RESULT 6			
Db	686	TTCTCTGTCAGACTGGAACCCCTGACCATCTGGTTCTACAGGAGAGTTACTTGGCG	745	ID	ADQ17814/c		
QY	221	PheSerThrGlnArgPheSerLeuAspLysLeuAspSerAlaIleHisLeuCysAsnAsn	240	AD	ADQ17814 standard; DNA; 101270 BP.		
Db	746	TTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGACAGCGCCATCCACCTGTGCACAAC	805	AC	ADQ17814;		
QY	241	AlaValGlnLysTyrLeuLysAsnAspValGlyArgSerProLeuLeuProAlaHisAsn	260	XX	26-AUG-2004 (first entry)		
Db	806	ACCGTCAGAAAGTACCTGAAGAATGATGTGGCGCGAGCCCTGCTGCCCGCACACAC	865	DT	Human soft tissue sarcoma-upregulated DNA - SEQ ID 631.		
QY	261	MetTrpThrSerThrArgPheGlnGluTyrLeuGlnArgGlnGlyArgGlyAlaValTrp	280	XX	soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;		
Db	866	ATGTGGACCAAGCAGGTTCCAGGAGTACTTCAGCGCCAGCGCGCTGCGCGTGTGG	925	XX	ds.		
QY	281	GlySerValIleTyrProSerMetLysAlaIleAlaHisAlaMetLysValAlaGln	300	OS	Homo sapiens.		
Db	926	GGCAGCGTCATCTACCGCTCATGAGAGGCGCATCGCCACGCCATGAAGGTGGCCACG	985	XX	WO2004048938-A2.		
QY	301	AspHisValGluProArgLysAsnSerPheGluLeuTyrGlyAlaAspPheValLeuGly	320	XX	10-JUN-2004.		
Db	986	GACCACGTGAGGCTTCGCAAGACAGCTTTGAGCTCTACGCGGCTGACTTCGTCCTTGGG	1045	XX	26-NOV-2003; 2003WO-US038193.		
QY	321	ArgAspPheArgProTrpLeuIleGluIleAsnSerSerProThrMetHisProSerThr	340	XX	26-NOV-2002; 2002US-0429739P.		
Db	1046	AGGACTTCAGGCCCTGGCTGATCGAGATCAATTCAGCCGCCACCATGCACCGTCCACG	1105	XX	(PROT-) PROTEIN DESIGN LABS INC.		
QY	341	ProValThrAlaGlnLeuCysAlaGlnValGlnGluAspThrIleLysValAlaValAsp	360	XX	Aziz N, Ginsburg WM, Zlotnik A;		
Db	1106	CCGGTCACGGCCAGCTGTGTGCACAGGTGCGAGGAGACCATCAAGGTGGCGTGGAC	1165	XX	WPI; 2004-441208/41.		
QY	361	ArgSerCysAspIleGlyAsnPheGluLeuLeuTrpArgGlnProValValGluProPro	380	XX	Early detection of soft tissue sarcoma comprises determining expression		
Db	1166	CGCAGCTGTGACATCGGCAACTTCGAGCTCTGTGGAGGCGCGGTGGTGGACCGCC	1225	XX	of a gene in a first soft tissue sample and a normal soft tissue sample		
QY	381	ProPheSerGlySerAspLeuCysValAlaGlyValSerValArgAlaArgArgGln	400	XX	PT and comparing the gene expression, also useful in treating soft tissue		
Db	1226	CCATTACGGGTCCGACCTCTGCTGGCGGGCGTCAGTGTGAGGAGAGCCAGGAGCGAG	1285	XX	sarcoma.		
				XX	Example 2; SEQ ID NO 631; 210pp; English.		
				XX	The invention relates to a novel method for detecting soft tissue sarcoma		
				XX	which comprises obtaining a first soft tissue sample from an individual,		
				XX	and a normal soft tissue sample from the same or different individual,		
				XX	determining the expression of a gene in both samples and comparing the		
				XX	expression of the gene in both soft tissue samples, where a higher level		

CC of protein expression in the first soft tissue sample indicates the
 CC presence of soft tissue sarcoma. The method of the invention has
 CC cyrostatic applications and may be useful for detecting soft tissue
 CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
 CC acid sequences may be useful in diagnostic and screening applications.
 CC The current sequence is that of a human soft tissue sarcoma-upregulated
 CC DNA of the invention. The current sequence is not shown within the
 CC specification per se but was submitted in CD format by the inventor.
 XX

SQ Sequence 101270 BP; 24151 A; 28044 C; 28100 G; 20975 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 4,62e-106 Length: 101270
 Score: 1802.00 Matches: 438
 Percent Similarity: 32.2% Conservative: 1
 Best Local Similarity: 32.1% Mismatches: 3
 Query Match: 63.1% Indels: 923
 DB: 12 Gaps: 3

US-10-635-977-2 (1-541) x ADQ17814 (1-101270)

QY 100 LeuValHisGlyAspAlaPheIleSerAsnSerArgAsnTyrPheSerGlnCysGlnAla 119
 DB 47222 CTTTGTAGTGGCGATGCTTTTCATCTCCATTCAGAAATATCTTTTGGCAGTGCAGGCT 47163
 QY 120 LeuLeuAsnArgIleThrSerValAsnProGlnThrAspIleAspGlyLeuArgAsnIle 139
 DB 47162 CTGCTGAATAGAAATCACGTCGTGAACCTCAGACGACATTGACGGGCTCGGAACATC 47103
 QY 140 TrpIleIleLysProAlaAlaLysSerArgGlyArg----- 151
 DB 47102 TGGATTATAAAGCCCGCGCCCAAGTCCCGGGCCG-AGGTGAGTCCCTGGTGGCGATGC 47044
 QY 151 ----- 151
 DB 47043 TCCTTGACCGGTTTCATAGCCACCAACGGCCACACTTGATGCCAGGAGCCGTGATGAG 46984
 QY 151 ----- 151
 DB 46983 GGACAGTTTCAGGCCCAACAGCACACAGGAAGGAAGGAGGAGCAGAACGCCGAGCA 46924
 QY 151 ----- 151
 DB 46923 GCGTGGGGTTCAGGACTCAGCGCTGGAGCAGAGCTGGAGGCGTCTTCACAGCATGTGT 46864
 QY 151 ----- 151
 DB 46863 TCCGATCACAGGCGAGGATCCGGTTGATGGTTAATCAGGATTTCTCCTTTGTATTCA 46804
 QY 151 ----- 151
 DB 46803 CAAACTGCTATGCTTTTCCTTTTATTTTATTTATTTATTTATTTATTTATTTATTTT 46744
 QY 151 ----- 151
 DB 46743 TGAACGAGTTTCACTCTTGTGTGCCAGGCTGAAGTGCAGTGGCGCAATCTTGCTCAC 46684
 QY 151 ----- 151
 DB 46683 TGCACCTCCACCTTCCAGGTTTAAGCAATTTCTCTGCTTAGCCTCCCAAGTAGCTGGG 46624
 QY 151 ----- 151
 DB 46623 ACTACAGCGGCCACCAACACCGGCTAATTTTGTATTTTATTTAGTAGAGATGAGTTTCA 46564
 QY 151 ----- 151
 DB 46563 CCATGTTGGCCAGGCTGGTCTTGAACCTCCGACCTCAAGTGATTCACCTGCGCTCGGCTC 46504
 QY 151 ----- 151
 DB 46503 CCAAGTCTGGGATTACAGGCATTAAGCCATAGCCCATCTCTCTGGCTCCCTTTTAA 46444

QY 151 ----- 151
 DB 46443 AGTGTCTCTGTCTGCTGCCCTTCACTGATCTCTGTGTCCCTCTGGGGTCTCAGAGGCTG 46384
 QY 151 ----- 151
 DB 46383 GTTCTGTTGGGGAGAGCCTGGGGGACTCAGGCACTGGGCTTTAGCCTCATCATGAGCGC 46324
 QY 151 ----- 151
 DB 46323 TAAGCTTCCCTGCTCCCCCATCGCCCCCACCACCAACCATGCTGCTGCTGGCCTTGGACAA 46264
 QY 151 ----- 151
 DB 46263 AGCCAGTGCAGGTCTCAGSGCCAAGTGGGACCTGCGCCAGGGCTCAGGAGCTCAGGTC 46204
 QY 152 -----AspIleValCys 155
 DB 46203 CCGCTGCACAGTGGCTGTGCATGTGGTGTGAGCGGTGGCCTCTTTCAGACATAGTGTGC 46144
 QY 156 MetAspArgValGluGluIleLeuGluLeuAlaAlaAspHisProLeuSerArgAsp 175
 DB 46143 ATGGACCGTGTGGAGAGATCTCTGGAGCTGGCAGCTGCAGACACCTCTTTCCAGGAC 46084
 QY 176 AsnLysTrpValValGlnLysTyrIleGluThrProLeuLeuIleCysAspThrLysPhe 195
 DB 46083 AACAAAGTGGTGGTCCAGAGATGATCATCGAGACGCCGCTGCTCATCTGTGACACCAAGTTC 46024
 QY 196 AspIleArgGlnTrpPheLeuValThrAspTrpAsnProLeuThrIleTrpPheTrpLys 215
 DB 46023 GACATCAGACAGTGGTCTCTCGTCAACGACTGGAAACCCCTGACCATCTGTGTTCTACAAG 45964
 QY 216 GluSerTyrLeuArgPheSerThrGlnArgPheSerLeuAspLysLeuAsp----- 232
 DB 45963 GAGATTACTTGGGTTTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGACAGGTGACGTG 45904
 QY 232 ----- 232
 DB 45903 CGGTGGGCGACCTGGACACCTCGGCGCAGGGAATGGGCTGCTGCTGTCATGCTGGAGCA 45844
 QY 232 ----- 232
 DB 45843 GCCGAGTGAAGAGCAGCACGCGGAGGCGCCCTCACTGCCACCTGCCACACGAGAGGC 45784
 QY 232 ----- 232
 DB 45783 ACAGATGGGGAACCTGAGGCTTGGAGGCTGATGATCTCCCGAGGGAAGGCCAGGCC 45724
 QY 232 ----- 232
 DB 45723 CACCGGCTTCAAAGCCAGCGCCCTACACCCACGCTCTGTGCATCATCTCTGGAAAATC 45664
 QY 232 ----- 232
 DB 45663 TTAAGAGSCACCAATTTTGTCTGTGCCAGGGGATTTGAATGGTGTGGCAGCAGCTGGG 45604
 QY 232 ----- 232
 DB 45603 CCCTGGGCGCAGGGGAGGCGGCTGTGGGCTGCTGTGGAGGCTCAGATATCCGCAAGTG 45544
 QY 232 ----- 232
 DB 45543 TGCTGCCAGGTTGATGGACCCCACTCACCCGGGTCCCGCGTGGGAGGAGGCGCA 45484
 QY 232 ----- 232
 DB 45483 GAGCCTGTGGAGTCTCAGGAGTGTGTGTGTCACACAGCAGCGCTGCCACCTCTGTCT 45424
 QY 232 ----- 232
 DB 45423 GCACCAAGGCTTCTCTGACTATTTGTGTGGCCCTTGTGTGTGAGCGCTAGGGGACC 45364
 QY 232 ----- 232

Dbb	45363	TGACGGCTGGGACAGGAAGGGGGTCTGCTTTTACGAGTGAGATCCAGGAGGSCCTGAGA	45304	Dbb	44283	TGGAGSACACCGATCCTGCTCTGTGTGACCATGTCTGGTCCGCTAGGGGCCAGGCGCTGCC	44224
Qy	232	-----	232	Qy	373	-----	373
Dbb	45303	GTACCGGACACAGACCTTGCAGGGAGTTAACTCCGGCCAGCGAGCGGTGCGGTCA	45244	Dbb	44223	CTGCTCTCCATGGGGCTCAGTGCCTCGTCTGGCA CCGAGAAAGGGGCTACTCCAGGGTG	44164
Qy	232	-----	232	Qy	373	-----	373
Dbb	45243	TGCCTGGTGGGTCACTGCTCGAAGTCA CACGAGAAGAGGCTTGGAGGAGCTGGCCAGGG	45184	Dbb	44163	CGGGGAGGGGCGCGCTCACTCAGGCTCAGTGCCTCTGTCCAGCTTTCTCTCCTCTTG	44104
Qy	232	-----	232	Qy	373	-----	373
Dbb	45183	GCTACGACCCTCACCTGCTCACTGCCTCTGCTCAAGTGGGCCCTGGAGGGGGTCCAGGAAC	45124	Dbb	44103	TTCCACGCTCACCCAGTCTCTGACTGTGTGTAGAGTGGGGCTCCATGCCCGCAGCCTC	44044
Qy	232	-----	232	Qy	373	-----	373
Dbb	45123	TCACCTGTGAAAGCAGCTCCACGAGCGGCAGTGAGAAGCCAGGGCCCCCTGCCAGCG	45064	Dbb	44043	CAGGCCACATATGCTGCTCTTTTGGCTTGGCTTGCAGTGCCTCTCOACTCTCAGGATCT	43984
Qy	232	-----	232	Qy	373	-----	373
Dbb	45063	CCCATTTCTGGGCTTCAGCTGTGAGCTGTGCGCTGGCCAAATCCATGGCCTCTCTGAG	45004	Dbb	43983	TTCCGGGGCTCCCTGTGCTGCGGATGTCTGGAGTCA TTGGGAAGCGTGGCCAATGCC	43924
Qy	232	-----	232	Qy	373	-----	373
Dbb	45003	CCCTGTTTCTCCATCTGGGGACAGTAACGCTCATCCAGCTGGGGCCCTCCAGCAGAGGT	44944	Dbb	43923	TCTCTCTTGGGGCTCTCTGTGGCTAACTCGGGAAAAAGACGCTAGTCCCTGGGTGT	43864
Qy	232	-----	232	Qy	373	-----	373
Dbb	44943	TGGAGTCTCAGGGGCGAGGTCTCTAGAGCCAGCTGGGTGGGCTCAGCAGCGCTGGAG	44884	Dbb	43863	GGGGCCCCCAGGCTGACCCCCAAGCCTCAGGAAGCCTTGGTTTCCCGCAGGTTCTCCAG	43804
Qy	232	-----	232	Qy	373	-----	373
Dbb	44883	GACCGGCTTTGGGTGCCCTCTTGGGCCAAGGAGGCTGAGCACCTGCCCCCTCGTCCCCC	44824	Dbb	43803	AGCCTGGGGCTCTCTGGGGGTCTCGTGGGACCCCGGAGTGTTCATGTTCACTCTCTGAG	43744
Qy	233	---SerAlaIleHisLeuCysAsnAlaValGlnIlySerLeuLysAsnAspValGly	251	Qy	373	-----	373
Dbb	44823	TGACGGCCATCACCTGTGCAACAGCGCGTCCAGAAGTACCTGAAGAATGATGGGC	44764	Dbb	43743	TCAGGTGGGGTCTGTGCCCCGACACGCCCTCTCTGCACAGACCCCTCTGCAGGGACCC	43684
Qy	252	ArgSerProLeuLeuProAlaHisAsnMetTrpThrSerThrArgPheGlnIlyrLeu	271	Qy	374	-----	374
Dbb	44763	CGAGCCCCCTGTGCCCGCACACAATGTGGACAGCACAGGTTCACAGGAGTACCTG	44704	Dbb	43683	CCCTATCCCGGCACAGCCACAGCACCCCCAACCTGTTCTTCCCAAGCGGTGTGTGA	43624
Qy	272	GlnArgGlnGlyArgGlyAlaValTrpGlySerValIleTrpProSerMetLysLysAla	291	Qy	378	uProProProPheSerGlySerAspLeuCysValAlaGlyValSerValArgArgAlaAr	398
Dbb	44703	CAGCGCCAGGCGCTGGCGCGTGTGGGGCAGCGTCACTACCGCTCCATGAAGAAGGCC	44644	Dbb	43623	GGCGCCCCCATTCAGCGGTCCGACCTCTGCGTGGCGGGGTCTCAGTGTGAGGAGGCCAG	43564
Qy	292	IleAlaHisAlaMetLysValAlaGlnAspHisValGluProArgLysAsnSerPheGlu	311	Qy	398	gArgGlnValLeuProValCysAsnLeuLysAlaSerAlaSerLeuLeuAspAlaGlnPr	418
Dbb	44643	ATCGCCACGCATGAAGGTGGCCAGGACACGCTGGAGCTTCGAAGAACAGCTTTGAG	44584	Dbb	43563	GAGGCAGGTGCTGCCGCTCTGCAACCTCAAGGCTCGGCTCTGTGGACGCGCAGCC	43504
Qy	312	LeuTrpGlyAlaAspPheValLeuGlyArgAspPheArgProTrpLeuIleGluIleAsn	331	Qy	418	oLeuLysAlaArgGlyProSerAlaMetProAspProAlaGlnGlyProProSerProAl	438
Dbb	44583	CTCTACGGGCTGACTTCGTCTTGGGAGGGACTTCAGGGCTTGGCTGATCGAGATCAAT	44524	Dbb	43503	GCTGAAGGCACGGGGCCCCCTCGGCATTCGCAGACCTTGCAGGACCCCGCATCACCAGC	43444
Qy	332	SerSerProThrMetHisProSerThrProValThrAlaGlnLeuCysAlaGlnValGln	351	Qy	438	aLeuGlnArgAspLeuGlyLeuLysGluGlyLysGlyLeuProLeuAlaLeuLeuAlaPr	458
Dbb	44523	TCCAGCCCCACCATGCACCCGCTCCACCGCGTCCAGCGCCAGCTGTGTGCACAGGTGAG	44464	Dbb	43443	TC1CCAGCGGACTTGGGACTGAGGAGAGAGGGGCTCCCCCTGGCTTGTGGCACC	43384
Qy	352	GluAspThrIleLysValAlaValAspArgSerCysAspIleGlyAsnPheGluLeuLeu	371	Qy	458	oLeuArgGlyValAlaAlaGluSerGlyGlyAlaAlaGlnProThrArgThrLysAlaAlaGl	478
Dbb	44463	GAGACACCATCAAGGTGGCGGTGGACCGCAGCTGTGACATCGGCAACTTCGAGCTCTG	44404	Dbb	43383	CTTAAGGGGGGACCGGAGCGGTGGAGCGGCACAGCCACCGCACAGCCCAACAGCTGTGG	43324
Qy	372	TrpArg-----	373	Qy	478	YLysValGluLeuLeuProAlaCysProCysArgHisValAspSerGlnAlaProAsnThrGl	498
Dbb	44403	TGGAGGCGAGGTGAGCCACGCCGCCCTCTGGGACTTTTGGTGGCGCGCCCATGTGGATG	44344	Dbb	43323	GAAGGTGGAGCTCCCGGCTCTGTCCTGTCAGTGGAGTGGAGTGGAGTGGAGTGGAGTGG	43264
Qy	373	-----	373	Qy	498	YValProValAlaGlnProAlaLysSerTrpAspProAsnGlnLeuAsnAlaHisProLe	518
Dbb	44343	CTTTGCCACCTGTGCTCACTGGGGGAGGGTGGAGCTGCAGCGCGCAGTGCAGCGGGTCTA	44284	Dbb	43263	TGTCCTTGGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAG	43204
Qy	373	-----	373	Qy	518	uGluProValLeuArgGlyLeuLysThrAlaGluGlyAlaLeuArgProProGlyGly	538
				Dbb	43203	GGAGCCTGTGTGCGGGGCTGAAGACAGACAGAGGGGCGCTGCTGCTCGCGCGCGAGG	43144

QY 373 ArgGln 374
 Db 721 AGGCAG 726

RESULT 8
 AAS99894
 ID AAS99894 standard; cDNA; 2380 BP.
 XX AC
 XX AAS99894;
 XX 12-MAR-2002 (first entry)
 DT
 XX Polynucleotide encoding human cytoskeleton-associated protein (CYSKP) #5.
 DE
 XX Human; cytoskeleton-associated protein; CYSKP; autoimmune disorder; ss;
 KW cell proliferative disorder; inflammatory disorder; prion disease;
 KW vesicle trafficking disorder; gastrointestinal disorder; muscle disorder;
 KW neurological disorder; cell motility disorder; reproductive disorder;
 KW spinal cord disease; central nervous system disorder; mental disorder;
 KW gene therapy; cancer.
 KW
 XX Homo sapiens.
 OS
 XX WO200185942-A2.
 PN
 XX 15-NOV-2001.
 XX
 XX 03-MAY-2001; 2001WO-US014355.
 PF
 XX 05-MAY-2000; 2000US-0201960P.
 PR
 XX 08-MAY-2000; 2000US-0202729P.
 PR
 XX 05-JUN-2000; 2000US-0209705P.
 PR
 XX 07-JUN-2000; 2000US-0210149P.
 PR
 XX 21-JUN-2000; 2000US-0213215P.
 PR
 XX (INCY-) INCYTE GENOMICS INC.
 PA
 XX Yue H, Tang YT, Au-Young J, Lu DAM, Baughn MR, Hillman JL;
 PI Azimzai Y, Lal P, Yao MG, Bandman O, Burford N, Batra S, Kearney L;
 PI Policky JL;
 XX
 DR WPI; 2002-062248/08.
 DR P-PSDB; AAU74334.
 XX
 XX New cytoskeleton-associated proteins and polynucleotides, useful for
 PT diagnosing, preventing and treating cell proliferative, autoimmune,
 PT inflammatory, neurological, cell motility, reproductive and muscle
 PT disorders.
 PS Claim 5; Page 171; 194pp; English.
 XX
 XX The invention relates to human cytoskeleton-associated polypeptides
 CC (CYSKP) and their associated polynucleotide sequences. The sequences are
 CC useful in the treatment of disorders associated with overexpression or
 CC underexpression of CYSKP in a patient. The disorders include cell
 CC proliferative disorders (such as cancer, actinic keratosis,
 CC arteriosclerosis, cirrhosis, hepatitis and psoriasis),
 CC autoimmune/inflammatory disorders (such as, asthma, atherosclerosis,
 CC osteoporosis, Crohn's disease, rheumatoid arthritis, diabetes mellitus
 CC and anaemia), vesicle trafficking disorders (such as
 CC hypercholesterolaemia, diabetes insipidus, Grave's disease and goitre),
 CC gastrointestinal disorders, prion diseases, neurological disorders (such
 CC as epilepsy, stroke, cerebral neoplasms, Alzheimer's disease,
 CC Huntington's disease, Parkinson's disease, amyotrophic lateral sclerosis
 CC and other motor neuron disorders), cell motility disorders, reproductive
 CC disorders (such as endometriosis and polycystic ovary syndrome), muscle
 CC diseases (such as myocarditis, migraine, hypertension, hypoglycaemia,
 CC myocardial infarction, epilepsy and muscular dystrophy), spinal cord
 CC diseases, central nervous system disorders (such as Down syndrome and
 CC cerebral palsy) and mental disorders (such as anxiety and schizophrenia).
 CC Sequences AAS99890-AAS99923 represent cDNA molecules encoding human CYSKP
 CC of the invention

XX SQ Sequence 2380 BP; 580 A; 685 C; 670 G; 445 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 2.67e-60 Length: 2380
 Score: 1070.50 Matches: 250
 Percent Similarity: 54.8% Conservative: 81
 Best Local Similarity: 41.4% Mismatches: 198
 Query Match: 37.5% Indels: 76
 DB: Gaps: 12

US-10-635-977-2 (1-541) x AAS99894 (1-2380)

QY 2 AlaSerSerIleuLeuYsrTrpValValSerHisGlnSerCysSerArgSerSerArgSer 21
 Db 285 GCCCGCAAGCTTCTCAAGCTGGTGGTGAAGTCTGAG-----TGGAGTCA 329
 QY 22 LysProArgAspGlnArgGluGluAlaGlySerSerAspLeuSerSerArgGlnAspAla 41
 Db 330 TACCCTATTTCAGGCAGTACAGGAAGAGGCTCAGGAGACAGCAGCCCAAGAAACAGGAG 389
 QY 42 GluAsnAlaGluAlaLysLeuArgGlyLeuProGlyGlnLeuValAspIleAlaCysLys 61
 Db 390 AAAAACCCAGTGTG-----GTGTCCCGAGAGTTTGTGATGAAGCTGTGTGT 437
 QY 62 ValCysGlnAlaTyrlLeuGlyGlnLeuGluHisGluAspIleAspThrSerAlaAspAla 81
 Db 438 GCGTGGGAGGAGTACCTTAGCAACTTGGCCCATCGACATCGAACAGGACCTGGAGGCC 497
 QY 82 ValGluAspLeuThrGluAlaGluTrpGluAspLeuThrGlnGlnTyrlSerLeuVal 101
 Db 498 CCGCTGTACTACACCCCGAGGCTGGTCCCTCTTCTCCAGCGTACTACCAAGTGGTC 557
 QY 102 HisGlyAspAlaPheIleSerAsnSerArgAsnTyrlPheSerGlnCysGlnAlaLeuLeu 121
 Db 558 CACGAAGGGGCGAAGCTCAGGCACCTCGACACTCAGGTCCAGCGCTGTGAGGACATCTG 617
 QY 122 AsnArgIleThrSerValAsnProGlnThrAspIleAspGlyLeuArgAsnIleTrpIle 141
 Db 618 CAGCAGCTCAGCGCGCTGGTGTACCCAGATAGACATGGAAGGGATCGCAACATCTGGATC 677
 QY 142 IleLysProAlaAlaLysSerArgGlyArgAspIleValCysMetAspArgValGluGlu 161
 Db 678 GTCAAGCCAGGAGCCCAAGTCCCTGGACGAGGATCATGTGATGGACCACTTGGAGGAG 737
 QY 162 IleLeuGluLeuAlaAlaAspHisProLeuSerArgAspAsnLysTrpValValGln 181
 Db 738 ATGCTGAAGCTGGTGAACCGCAACCCCGTGGTGTATGAAGAGCGCAAGTGGTGGTCAG 797
 QY 182 LysTyrlIleGluThrProLeuLeuIleCysAspThrLysPheAspIleArgGlnTrpPhe 201
 Db 798 AAGTATATTGAGCGGCCCCCTCTCATCTTTGGGCCCAAGTTTGACCTCAGACAGTGGTTC 857
 QY 202 LeuValThrAspTrpAsnProLeuThrIleTrpPheTyrlLysGluSerTyrlLeuArgPhe 221
 Db 858 CTGGTAACCTGACCTGGAAACCCACTTACGCTGGTGTCTACCGCGACAGCTATATCCGCTTT 917
 QY 222 SerThrGlnArgPheSerLeuAspLysLeuAspSerAlaIleHisLeuCysAsnAsnAla 241
 Db 918 TCCAGCGAGCCCTTCTCCCTGAAGAACCTGCAACACTCAGTGCACCTGTGTGCAACAACCTC 977
 QY 242 ValGlnLysTyrlLeuLysAsnAspValGlyArgSerProLeuLeuProAlaHisAsnMet 261
 Db 978 ATCCGAAGACCTTGGAGAACTCATGCCATCGGATCCACTGCTTCGCCGACACACATG 1037
 QY 262 TrpThrSerThrArgPheGlnGluTyrlLeuGlnArgGlnGlyAlaValTrpGly 281
 Db 1038 TGGTCTAGCAGAGGTTCCAGGCCCACTGCAGGAGATGGGTGCCCAAAATGTGTGTC 1097
 QY 282 SerValIleTyrlProSerMetLysLysAlaIleAlaHisLalaMetLysValAlaGlnAsp 301
 Db 1098 ACCATCATCGTGGCTGAGGATGCTGTGATCCACGCACTTCAGACCTCCAGGAC 1157

Qy	302	HisValGluProArgLysAsnSerPheGluLeuTyrglyAlaAlaPheValLeuGlyArg	321
Db	1158	ACCGTGCAGTTCGGAAGCCAGCTTTGAGCTCTATGCGCTGACTTCGTGTTCCGGAG	1217
Qy	322	AspPheArgProTtpLeuIleGluLeuAsnSerSerProThrMetHisProSerThrPro	341
Db	1218	GACTTCAGCCCTCGCTGATTGAGATCAACGCCAGCCCCACGATGGCACCCTCCACAGCA	1277
Qy	342	ValThrAlaGlnLeuCysAlaGlnValGlnIleAspThrIleLysValAlaVal-----	359
Db	1278	GTCACTGCCCGGCTCTGTGCTGGCGTGCAAGCTGACACCCCTGCCGCGTGGTCATTGACCGG	1337
Qy	360	-----AspArgSerCysAspIleGlyAsnPheGluLeuLeuTrpArgGlnProValVal	377
Db	1338	ATGCTGGACCGCAACTGTGACACAGGAGCCTTTGAGCTCATCTATAGCAGCCTGTGTGTG	1397
Qy	378	GluProProPheSerGlySerAspLeuCysValAlaGlyValSerValArgArg---	396
Db	1398	GAGGTGCTCAATATGTGGGCATCCGGCTCTGTGTAGAGGGCTTCACCATCAAGAAGCCC	1457
Qy	397	-----AlaArgArgGlnValLeuProValCysAsnLeu-----	407
Db	1458	ATGGCGATGTGTCATCGCGGATGGGGGTCCGCCACGACGTCCCTCTGTGACCCAGCGA	1517
Qy	408	-----LysAlaSer-AlaSerLeuLeuAspAlaGlnProLeu---LysAl	421
Db	1518	GGCTCTGGGGAAGCAAGGACTCGGGGATCCCTACCCACAGGTCAGCTCTAGGAAGGC	1577
Qy	421	aArgGlyProSerAlaMetProAspProAlaGlnGlyProProSerProAlaLeuGln--	440
Db	1578	ACTGGGCCAGAGCCTTGGGGCACAGTGAAGACGAGTCTCCACTGCGCACCACTTCAGCC	1637
Qy	441	-ArgAspLeuGlyLeuLysGluGluLysGlyLeuProLeuAlaLeuAlaProLeuAr	460
Db	1638	CCCGAAGGGGAAGAAAGGCAAGGCCAAAAGGGCCACAGCCCTGGTCTGCCCAATCTC	1697
Qy	460	gGlyAlaAla-----GluSerGlyGlyAlaAlaGlnProThrArgThrLysAlaAl	477
Db	1698	TGGGAGTGGGATGCCGCCAGCACAGGATGGGCTGCATTTTC-ACCATGACCTTTTCTAG	1756
Qy	477	aglyLysValGlu-----LeuProAlaCysProCysArgHisValAspSerGlnAl	494
Db	1757	TGGGGAAGCAACCCACCACCTTGAACAGATTGCCACTGAGTCCGAAGAACCCCGAGC	1816
Qy	494	aProAsnThrGlyValProValAlaGln-----ProAlaLys	506
Db	1817	CCTGGGTAAAGCCATTCCTCCCAAAACCCGAGTGTTCNAGCGGATTTATTCCTGCTCT	1876
Qy	506	sSerTrpAspProAsnGlnLeuAsnAlaHisPro-----	517
Db	1877	CCAGGCCCTCTCCCAACCACTGATCAGCCACCCCAAGAGCCACCCAGTAGCAAGTA	1936
Qy	517	-----	517
Db	1937	AAAGCCACTACTCACAAGATTTGTTTAAATAFACAGCCAAATTAGTGGGCAAGGTG	1996
Qy	518	-LeuGluProValLeuArgGlyLeuLysThrAlaGluGlyAlaLeuArgProProGln	537
Db	1997	GTGTGAGCCTGTGTGTCAGGCTACTCGGAGGCTATAGGAT---CGTTGAGCCCGGG	2053
Qy	537	yGlyLysGly	540
Db	2054	AGGTCAAGGC	2063

RESULT 9

AAH16735

ID ID AAH16735 standard; cDNA; 2326 BP.

XX AC

XX AAH16735;

DT 26-JUN-2001 (first entry)

XX Human cDNA sequence SEQ ID NO:15920.

XX	Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
KW	
XX	
OS	Homo sapiens.
XX	
XX	
PN	EP1074517-A2.
XX	
XX	
PD	07-FEB-2001.
XX	
XX	
PF	28-JUL-2000; 2000EP-00116126.
XX	
XX	
PR	29-JUL-1999; 99JP-00248036.
PR	27-AUG-1999; 99JP-00300253.
PR	11-JAN-2000; 2000JP-00118776.
PR	02-MAY-2000; 2000JP-00183767.
PR	09-JUN-2000; 2000JP-00241899.
XX	
XX	
PA	(HELI-) HELIX RES INST.
XX	
PI	Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI	Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX	
XX	
DR	WPI; 2001-318749/34.
XX	
XX	
PT	Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT	length cDNAs defined in the specification, and for the detection and/or
PT	diagnosis of the abnormality of the proteins encoded by the full-length
PT	cDNAs.
XX	
PS	Claim 8; SEQ ID NO 15920; 2537bp + Sequence Listing; English.
XX	

The present invention describes primer sets for synthesising 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesising polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

Sequence 2326 BP; 524 A; 701 C; 642 G; 459 T; 0 U; 0 Other;

Alignment Scores:		
Pred. No.:	3.08e-59	Length:
Score:	1054.00	Matches:
Percent Similarity:	55.4%	Conservative:
Best Local Similarity:	41.4%	Mismatches:
Query Match:	36.5%	Indels:
DB:	4	Gaps:
		2326
		233
		79
		166
		86
		12

US-10-635-977-2 (1-541) x AAH16735 (1-2326)

Qy	2	AlaSerSerIleLeuLysTrpValValSerHisGlnSerCysSerArgSerArgSer	21
Db	544	GGCCCAACGGTTCTCAAGCTGGTGGTGAAGTCTGAG	588
Db		GGCCCAACGGTTCTCAAGCTGGTGGTGAAGTCTGAG	588
Qy	22	LysProArgAspGlnArgGluGluAlaGlySerSerAspLeuSerSerArgGlnAspAla	41
Db	589	TACCCTATTTCAGGCAGTAGAGGAGAGAGCCCTCAGGACCAAGCAGCCCAAGAACAGGAG	648

The invention relates to a new isolated polynucleotide (a Human kidney tumour specific cDNA) comprising any one of the 1855 sequences identified in the specification (for their complements, degenerate variants, sequences consisting of at least 20 contiguous residues them, sequences that hybridise to them under highly stringent conditions or sequences having at least 75 or 90% sequence identity to the 1855 sequences. Also included are detecting/determining the presence of cancer in a patient, stimulating an immune response in a patient; treating kidney cancer in a patient, an isolated polypeptide encoded by one of the 1855 sequences, an expression vector comprising the polynucleotide operably linked to an expression control sequence, a host cell transformed/transfected with the vector, an isolated antibody (or its antigen-binding fragment) that specifically binds to the protein, a fusion protein comprising at least one the proteins, stimulating and/or expanding T-cells specific for a tumour protein, an isolated T-cell population comprising the T-cells, a composition comprising a first component (such as a carrier or immunostimulant) and a second component (comprising one of the polynucleotides, the polypeptides, an antibody, T-cell or an antigen-presenting cell that expresses the polynucleotide) and a diagnostic kit comprising at least one of the oligonucleotides, or at least one antibody and a detection reagent comprising a reporter group. The polynucleotides, polypeptides, antibodies and antigen-presenting cells are useful for detecting the presence of, and treating cancer, particularly kidney cancer by stimulating and/or expanding T-cells specific for a tumour protein, and stimulating immune response in a patient. The present sequence is one of the Human kidney tumour specific cDNAs. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from USPTO at seqdata.uspto.gov/sequence.html?DocID=20030109434.

Sequence 2538 BP; 555 A; 717 C; 734 G; 532 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	5,59e-57	Length:	2538
Score:	1020.00	Matches:	220
Percent Similarity:	55.5%	Conservative:	69
Best Local Similarity:	42.2%	Mismatches:	151
Query Match:	35.7%	Indels:	81
DB:	7	Gaps:	12

US-10-635-977-2 (1-541) X ADS73168 (1-2538)

Qy	54	GlnLeuValAspIleAlaCysLysValCysGlnAlaTyrLeuGlyGlnLeuGluHisGlu	73
Db	835	GAGTTTGTGATGAAGCTCTGTGTGCGAGGAGTACCTTAGCAACTGGCCACATG	894
Qy	74	AspIleAspThrSerAlaAspAlaValGluAspLeuThrGluAlaGluTrpGluAspLeu	93
Db	895	GACATGCACAAGGACCTGGAGGCCCTGTACCTCCCCCGAGGCGCTGTCCCTCTTC	954
Qy	94	ThrGlnGlnTyrTrpSerLeuValHisGlyAspAlaPheIleSerAsnSerArgAsnTyr	113
Db	955	CTCCAGCGCTACTACCAAGTGTCCACGAGGGGCGAGAACTCAGGCACCTCGACACTCAG	1014
Qy	114	PhaSerGlnCysGlnAlaLeuLeuAsnArgIleThrSerValAsnProGlnThrAspIle	133
Db	1015	GTCACGCGTGTGAGGACATCTCGACAGCATGCGAGCCGTGTGTACCCAGATAGATG	1074
Qy	134	AspGlyLeuArgAsnIleTrpIleLeuProAlaIalysSerArgGlyArgAspIle	153
Db	1075	GAAGGGGATCGCAACATCTGGATCTGGTGAACCGAGGACCAAGTCCCGGACAGGGCATC	1134
Qy	154	ValCysMetAspArgValGluGluIleLeuGluLeuAlaIalAspHisProLeuSer	173
Db	1135	ATGTGATGACCACTCGAGGAGATGCTGAAGCTGTGAACGCAACCCCGGTGTGATG	1194
Qy	174	ArgAspAsnLysTrpValValGlnLysTyrIleGluThrProLeuLeuIleCysAspThr	193
Db	1195	AAGACGGCAAGTGGTGGTGGCAAGATATTGAGCGGCCCTCTCTCATCTTTGGCACC	1254
Qy	194	LysPheAspIleArgGlnTrpPheLeuValThrAspTrpAsnProLeuThrIleTrpPhe	213
Db	1255	AGATTTGACCTCAGACAGTGGTCTCTGGTAACTGACTGGAACCACTTACCGGTGGTTC	1314

QY 538 Y 538
Db 2260 A 2260
RESULT 11
ADW42022
ID ADW42022 standard; cDNA; 2538 BP.
XX
XX
AC ADW42022;
XX
XX 24-MAR-2005 (first entry)
XX cDNA elevated in kidney tumor cells SEQ ID NO 1765.
XX gene therapy; diagnosis; pharmaceutical; cancer; cytostatic;
XX genitourinary disease; neoplasm; renal tumor; gene; ss.
XX Unidentified.
XX WO200274237-A2.
XX
XX 26-SEP-2002.
XX
XX 19-MAR-2002; 2002WO-US010055.
XX
XX 19-MAR-2001; 2001US-0277245P.
XX 21-DEC-2001; 2001US-0343340P.
XX (CORI-) CORIXA CORP.
XX
XX
XX Algate PA, Mannion J, Gaiger A, Gordon B, Harlocker SL;
XX WPI; 2002-759855/82.
XX
XX New isolated polynucleotides and polypeptides, useful for detecting the
XX presence of, and treating cancer, particularly kidney cancer by
XX stimulating T-cells specific for a tumor protein, and stimulating immune
XX response in a patient.
XX
XX Claim 13; SEQ ID NO 1765; 252pp; English.
XX
XX The invention describes a new isolated polynucleotide (I) comprising: any
XX one of the 1855 sequences identified in the specification; complements or
XX degenerate variants of (a); sequences consisting of at least 20
XX contiguous residues of (a); sequences that hybridize to (a) under highly
XX stringent conditions; or sequences having at least 75 or 90% sequence
XX identity to (a). Also described are: detecting (M1) or determining the
XX presence of cancer in a patient; stimulating (M2) an immune response in a
XX patient; treating (M3) kidney cancer in a patient; an isolated
XX polypeptide (II) encoded by (I) and comprising, or having at least 70 or
XX 90% sequence identity to, any one of the 8 sequences identified in the
XX specification; an expression vector (III) comprising (I) operably linked
XX to an expression control sequence; a host cell (IV) transformed or
XX transfected with (III); an isolated antibody (V) or its antigen-binding
XX fragment that specifically binds to (II); a fusion protein (VI)
XX comprising at least one (II); an oligonucleotide (VII) that hybridizes to
XX the nucleotide sequences cited above under highly stringent conditions;
XX stimulating (M4) and/or expanding T-cells specific for a tumor protein;
XX an isolated T-cell population (VIII) comprising the T-cells in (M4); a
XX composition (IX) comprising a first component such as a carrier or
XX immunostimulant and a second component comprising (I), the polypeptide
XX encoded by (I), an antibody or its antigen-binding fragment that
XX specifically binds to (II), (VI), or an antigen-presenting cell that
XX expresses the polynucleotide; and a diagnostic kit (X) comprising at
XX least one of the oligonucleotide, or at least one antibody and a
XX detection reagent comprising a reporter group. The polynucleotides,
XX polypeptides, antibodies and antigen-presenting cells are useful for
XX detecting the presence of, and treating cancer, particularly kidney
XX cancer by stimulating and/or expanding T-cells specific for a tumor
XX protein, and stimulating immune response in a patient. This sequence
XX represents a kidney tumour cDNA, expression of which is increased in
XX kidney tumors.

SQ Sequence 2538 BP; 555 A; 717 C; 734 G; 532 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 5.59e-57 Length: 2538
Score: 1020.00 Matches: 220
Percent Similarity: 55.5% Conservative: 69
Best Local Similarity: 42.2% Mismatches: 151
Query Match: 35.7% Indels: 81
DB: 7 Gaps: 12
US-10-635-977-2 (1-541) x ADM42022 (1-2538)
QY 54 GlnLeuValAspIleAlaCysLysValCysGlnAlaTyrLeuGlyGlnLeuGluHisGlu 73
Db 835 GAGTTTGTGATGAAGCTCTGTGCGTGGAGGAGTACCTTAGCAACTTTGGCCACATG 894
QY 74 AspIleAspThrSerAlaAspAlaValGluAspLeuThrGluAlaGluTrpGluAspLeu 93
Db 895 GACATCGACAAGGACCTGGAGGCCCGCTGTACCTCACCCCGAGGGCTGGTCCCTCTTC 954
QY 94 ThrGlnGlnTyrTyrSerLeuValHisGlyAspAlaPheIleSerAsnSerArgAsnTyr 113
Db 955 CTCACGCGCTACTACCAAGTGTGTCCAGAAAGGGGCGAGAACTCAGGCACCTCGACACTCAG 1014
QY 114 PheSerGlnCysGlnAlaLeuLeuAsnArgIleThrSerValAsnProGlnThrAspIle 133
Db 1015 GTCCACGCGTGTGAGGACATCTGCAGCAGCTGCAGGCGGTGGTACCCCGAGATAGACATG 1074
QY 134 AspGlyLeuArgAsnIleTrpIleLeuLysProAlaAlaLysSerArgGlyArgAspIle 153
Db 1075 GAAGGGGATCGCAACATCTGGATCGTGAAGCCAGGAGCCAAAGTCCCGCGACGAGGCATC 1134
QY 154 ValCysMetAspArgValGluGluLeuLeuGluAlaAlaAspHisProLeuSer 173
Db 1135 ATGTGATGCACACCTCGAGGAGATGCTGAAGCTGGTGAACCGGCAACCCCGTGGTGATG 1194
QY 174 ArgAspAsnLysTrpValValGlnLysTyrIleGluThrProLeuLeuLysCysAspThr 193
Db 1195 AAGGACGGCAAGTGGTGGTGCAGAGTATATTGACGGGCCCTCTCTCATCTTTGGCACC 1254
QY 194 LysPheAspIleArgGlnTrpPheLeuValThrAspTrpAsnProLeuThrIleTrpPhe 213
Db 1255 AAGTTTGACTCAGACAGTGGTTCTCTGGTAACTGACTGGAACCCCACTTACCGTGTGGTTC 1314
QY 214 TyrLysGluSerTyrLeuArgPheSerThrGlnArgPheSerLeuAspLysLeuAspSer 233
Db 1315 TACCGGACAGCTATATCCGCTTTCCAGCGACCCCTTCTCCCTGAAGAACCTGGGACAC 1374
QY 234 AlaIleHisLeuCysAsnAsnAlaValGlnLysTyrLeuLysAsnAspValGlyArgSer 253
Db 1375 TCAGTGACCTGTGCAACAACTCCATCCAGAGCACTGGAGAACTCATGCCATCGGCAT 1434
QY 254 ProLeuLeuProAlaHisAsnMetTrpThrSerThrArgPheGlnGluTyrLeuGlnArg 273
Db 1435 CCATGCTTCCCGCAGACACATGTGTCTAGCCAGAGGTTCCAGGCGCCACCTGCAGGAG 1494
QY 274 GlnGlyArgGlyAlaValTrpGlySerValIleTyrProSerMetLysLysAlaIleAla 293
Db 1495 ATGGTGCCGCCAAATGCTTGGTCCACCATCATCGTCCCTGGCATGAAGGATGTGTGATC 1554
QY 294 HisAlaMetLysValAlaGlnAspHisValGluProArgLysAsnSerPheGluLeuTyr 313
Db 1555 CAGCATTTCAGACCTCCAGGACACCGTGCAGTGTGCGAAGGCCAGCTTTGAGTCTAT 1614
QY 314 GlyAlaAspPheValLeuGlyArgAspPheArgProTrpLeuLeuGluIleAsnSerSer 333
Db 1615 GCGCTGACTTGTGTTCCGGGGAGGACTTCCAGCCCTGGCTGATTGATGATCAACGCCAGC 1674
QY 334 ProThrMetHisProSerThrProValThrAlaGlnLeuCysAlaGlnValGlnGluAsp 353
Db 1675 CCCACGATGCACCCCTCCAGCAGTCACTGCCCGGCTCTGTGTGCGGCTCAAGCTGAC 1734
QY 354 ThrIleLysValAlaVal-----AspArgSerCysAspIleGlyAsnPheGlu 369

```
Db 1735 ACCTGCGGTGTTCAATGACCGGATGCTGGACCGCACTGTGACACGAGCCCTTTGAG 1794
Qy LeuLeuTrpArgGlnProVal---ValGluPro-ProProPheSerGlySerAspLeuCy 388
Db 1795 CTCATCTATAAGCAGCCGCTCACCACTTCCCGAGCTCCACCAAGGC-----1843
Qy 388 sValAlaGlyValSerValArgAlaArgGlnValLeuProValCysAsnLeuLy 408
Db 1843 -----1843
Qy 408 sAlaSerAlaSerLeuLeuAspAlaGlnProLeuLyAlaArgGlyProSerAlaMetPr 428
Db 1844 ----CCAGTGGCTTCTCCCACTGACTCCGACACGAGGCCAGGTCTTCACAGCAGCAGC 1899
Qy 428 oAspProAlaGlnGlyProProSerProAlaLeuGlnArgAspLeuGlyLeuLyGluG1 448
Db 1900 ACAGCAAGCTGTGGGCACTAAGGCCCTGTGCACACAGCAAA---GGCTTTGAGGACTC 1956
Qy 448 uLys-----GlyLeuProLeuAlaLeuAlaPro----LeuArgGlyAla-----462
Db 1957 TACCACGCGCTAAGGCTTTCATTCTCCCTCCCAACCACTTGATTTCAAGGTGGCACCCA 2016
Qy 463 -----AlaGluSerGlyValAlaAlaGlnProThrArgThr-----474
Db 2017 GCATCCTGAAGCCAAAGAGGTGGGCTCGACTGTGACTACACCCAGTGGACAGTGCT 2076
Qy 475 -----LybAlaAlaGlyLybValGluLeuProAlaCysProCysAr 488
Db 2077 GAGCAGCGGTTCAGGCTGGAGGCGACAGGAGGAGGAGTCCAGGCTGG-----2128
Qy 488 gHisValAspSerGlnAlaProAsnThrGlyValProValAlaGlnProAlaLysSerTr 508
Db 2129 -----CTGGCACCACCAAGGAGGAGAGCTG 2151
Qy 508 pAspProAsnGlnLeuAsn-----AlaHisProLeuGluProValLe 522
Db 2152 GTCTCCCTCAGAAGCCCTTCTCTCCACAGACTTCTGATCATCTTCTCCCTCTCT 2211
Qy 522 uArgGlyLeuLybThrAlaGluGlyValAlaLeuArgPro-----ProProGlyG1 538
Db 2212 T-----TCACACCGAGGCTCTGCTCTCTGCTGCTCGAGGCCCCAGCTGG 2259
Qy 538 Y 538
Db 2260 A 2260

RESULT 12
ACN37881
ID ACN37881 standard; cDNA; 2553 BP.
XX
AC ACN37881;
XX
DT 18-NOV-2004 (first entry)
XX
DE Tumour-associated antigenic target (TAT) cDNA DNa324273, SEQ ID NO:1057.
XX
KW Tumour-associated antigenic target; TAT; human; overexpression; cancer;
KW tumour; diagnosis; cell proliferative disorder; breast cancer;
KW colorectal cancer; lung cancer; ovarian cancer; liver cancer;
KW central nervous system cancer; bladder cancer; pancreatic cancer;
KW cervical cancer; melanoma; leukaemia; hybridisation probe;
KW chromosome identification; chromosome mapping; gene mapping;
KW gene therapy; cytostatic; gene; ss.
XX
OS Homo sapiens.
XX
FN WO2004030615-A2.
XX
PD 15-APR-2004.
XX
PF 29-SEP-2003; 2003WO-US028547.
XX
```

PR 02-OCT-2002; 2002US-0414971P.

(GETH) GENENTECH INC.

Wu TD, Zhang Z, Zhou Y;

WPI; 2004-347921/32.

P-PSDB; ABM80420.

New tumor-associated antigenic target polypeptides and nucleic acids, useful in preparing a medicament for treating or detecting a proliferative disorder, e.g. breast, lung, colorectal, ovarian or prostate cancer or tumor.

Claim 1; SEQ ID NO 1057; 7273pp; English.

The invention relates to human tumour-associated antigenic target (TAT) polypeptides, and their related nucleic acids. The TAT polypeptides are overexpressed in cancer tissues compared to normal tissues, and may thus serve as effective targets for the diagnosis and treatment of cancer in mammals. The invention also relates to nucleic acid and polypeptide sequences at least 80% identical to the TAT nucleic acids and polypeptides; expression vectors and host cells comprising a TAT nucleic acid; an antibody specific for a TAT polypeptide; a peptide or organic molecule which binds to a TAT polypeptide; fusion proteins comprising a TAT polypeptide; and methods and compositions for the treatment or diagnosis of cancer in mammals. TAT polypeptides, nucleic acids, antibodies, antagonists, binding molecules and compositions are useful for diagnosing or treating a cell proliferative disorder associated with increased TAT expression, particularly cancers such as breast cancer, colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder cancer, pancreatic cancer, cervical cancer, cancers of the central nervous system, melanoma and leukaemia. TAT nucleic acids may further be used as hybridisation probes, in chromosome and gene mapping, in chromosome identification and in gene therapy. The present sequence represents a TAT nucleic acid of the invention

SQ Sequence 2553 BP; 575 A; 716 C; 733 G; 529 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	5,63e-57	Length:	2553
Score:	1020.00	Matches:	220
Percent Similarity:	55.5%	Conservative:	69
Best Local Similarity:	42.2%	Mismatches:	151
Query Match:	35.7%	Indels:	81
DB:	13	Gaps:	12

US-10-635-977-2 (1-541) x ACN37881 (1-2553)

Qy	54	GlnLeuValAspIleAlaCysLysValCysGlnAlaTyrLeuGlyGlnLeuGluHisGlu	73
Db	835	GAGTTTGTGGATGAAGCTCTGTGTCGTCGAGGAGTACCTTAGCACTTGGCCACATG	894
Qy	74	AspIleAspThrSerAlaAspAlaValGluAspLeuThrGluAlaGluTrpGluAspLeu	93
Db	895	GACATGCACAGGACCTGGAGGCCCGCTGTACCTCACCCCGAGGGCTGGTCTCTCTTC	954
Qy	94	ThrGlnGlnTyrTyrSerLeuValHisGlyAspAlaPheIleSerAsnSerArgAsnTyr	113
Db	955	CTCAGCGCTACTACCAAGTGGTCCAGAGGGCGCAACCTCAGGCACCTCGACACTCAG	1014
Qy	114	PheSerGlnCysGlnAlaLeuLeuAsnArgIleThrSerValAsnProGlnThrAspIle	133
Db	1015	GTCCAGCGCTGTGAGGACATCCTGCAGCAGCTGCAGGCGGTGTATCCCAATAGACATG	1074
Qy	134	AspGlyLeuArgAsnIleTrpIleLysProAlaLysSerArgGlyArgAspIle	153
Db	1075	GAAGGGATCGCAACATCTGGATCGTGAAGCAGCAGGACCAAGTCCCGCGGACGAGCATC	1134
Qy	154	ValCysMetAspArgValGluGluIleuGluLeuAlaAlaAspHisProLeuSer	173
Db	1135	ATGTGCATGGACCACCTGGAGGAGATGCTGAAGCTGTGTGAACGCAACCCCGTGTGATG	1194

QY 174 ArgAspAsnLysTrpValValGlnLysTyrIleGluThrProLeuLeuIleCysAspThr 193
 : : : : :
 Db 1195 AAGGACGGCAAGTGGTGGCAGAAAGTATTATGAGCGGCCCTCTCTCATCTTTGGCACC 1254
 : : : : :
 QY 194 LysPheAspIleArgGlnTrpPheLeuValThrAspTrpAsnProLeuThrIleTrpPhe 213
 : : : : :
 Db 1255 AGTTTGACTCAGACAGTGGTCTCTGGTAACAGTACTGGAACCCACTTACCGGTGGTTC 1314
 : : : : :
 QY 214 TyrLysGluSerTyrLeuArgPheSerThrGlnArgPheSerLeuAspLysLeuAspSer 233
 : : : : :
 Db 1315 TACCGCGACAGCTATATCGCTTTTCACGACGCCCTTCTCCCTGAAGAACCTGGACAAC 1374
 : : : : :
 QY 234 AlaIleHisLeuCysAsnAsnAlaValGlnLysTyrLeuLysAsnAspValGlyArgSer 253
 : : : : :
 Db 1375 TCAGTCACCTGTGCACAACTCCATCCAGAACACCTGGAGAACTCATGCCATCGGCAT 1434
 : : : : :
 QY 254 ProLeuProAlaHisAsnMetTrpThrSerThrArgPheGlnGluThrValGlnArg 273
 : : : : :
 Db 1435 CCACTGCTTCGGCCACACACATGGTGTAGCCAGAGGTTCCAGGCCCACTTCGAGGAG 1494
 : : : : :
 QY 274 GlnGlyArgGlyAlaValTrpGlySerValIleTyrProSerMetLysLysAlaIleAla 293
 : : : : :
 Db 1495 ATGGGTGCCCAAAATGCTTGGTCCACCATCATCGTGCCTGATGAAGGATGCTGTGATC 1554
 : : : : :
 QY 294 HisAlaMetLysValAlaGlnAspHisValGluProArgLysAsnSerPheGluLeuTyr 313
 : : : : :
 Db 1555 CACGCATCTCAGACCTCCAGGACACCGTGCAGTGTGGGAAGCCAGCTTTGAGCTCTAT 1614
 : : : : :
 QY 314 GlyAlaAspPheValLeuGlyArgAspPheArgProTrpLeuIleGluIleAsnSerSer 333
 : : : : :
 Db 1615 GGGCGTACTCTGTTTCGGGGAGACTTCCAGCCCTGGCTGATTGAGATCAACGCCAGC 1674
 : : : : :
 QY 334 ProThrMetHisProSerThrProValThrAlaGlnLeuCysAlaGlnValGlnGluAsp 353
 : : : : :
 Db 1675 CCCACGATGCGACCCCTCCACAGCAGTCACTGCGCGCTCTGTGTCGGCGTGAAGCTGAC 1734
 : : : : :
 QY 354 ThrIleLysValAlaVal-----AspArgSerCysAspIleGlyAsnPheGlu 369
 : : : : :
 Db 1735 ACCCTGCGCTGTGTCATTGACCGGATGCTGGACCGCAACTGTGCACAGAGGCTTTGAG 1794
 : : : : :
 QY 370 LeuLeuTrpArgGlnProVal---ValGluPro-ProProPheSerGlySerAspLeuCy 388
 : : : : :
 Db 1795 CTATCTATTAAGCAGCCCGCTCACACTTCCCGAGCTCCACCAAGC-----1843
 : : : : :
 QY 388 sValAlaGlyValSerValArgArgAlaArgGlnValLeuProValCysAsnLeuLy 408
 : : : : :
 Db 1843 -----1843
 : : : : :
 QY 408 sAlaSerAlaSerLeuLeuAspAlaGlnProLeuLysAlaArgGlyProSerAlaMetPr 428
 : : : : :
 Db 1844 ---CCAGCTGCTTCTCCCATGTACTCCGACACACAGGGCCAGGTCTCAGACGACAGC 1899
 : : : : :
 QY 428 oAspProAlaGlnGlyProProSerProAlaLeuGlnArgAspLeuGlyLeuLysGluL 448
 : : : : :
 Db 1900 ACAGCAAGCTGGTGGCAGCTAAGGCCCTGTGCACACAGGAA---GGCTTTAGGACTC 1956
 : : : : :
 QY 448 uLys-----GlyLeuProLeuAlaLeuLeuAlaPro---LeuArgGlyAla-----462
 : : : : :
 Db 1957 TACCCACGGCTAAGTCTTCTATTTCCTCCACCGAACCTTGATTTCAGAGTGGCACCCA 2016
 : : : : :
 QY 463 -----AlaGluSerGlyAlaAlaGlnProThrArgThr-----474
 : : : : :
 Db 2017 GCATCTGTAAGCAAGAAAGTGGCGCTCGACCTGTGACTCACACCCAGTGGACAGTGCT 2076
 : : : : :
 QY 475 -----LysAlaAlaGlyLysValGluLeuProAlaCysProCysAr 488
 : : : : :
 Db 2077 GAGCAGGGGTACGGCTGGAGGCCAGGACGAGGCGAGCTCCAGGCTGG-----2128
 : : : : :
 QY 488 gHisValAspSerGlnAlaProAsnThrGlyValProValAlaGlnProAlaLysSerTr 508
 : : : : :
 Db 2129 -----CTGCGACCCCAAGGAAGAGCTG 2151
 : : : : :
 QY 508 pAspProAsnGlnLeuAsn-----AlaHisProLeuGluProValLe 522
 : : : : :

Db 2152 GTCTCCTCCAGAGGCCCTTCTCCACAGACTTCTGATCATCTCCTCTTCTCCCTCCT 2211
 : : : : :
 QY 522 uArgGlyLeuLysThrAlaGluGlyAlaLeuArgPro-----ProProGlyGI 538
 : : : : :
 Db 2212 T-----TCACACGAGGCTCTCTCTCTGTGCTCCGAGGCCCCCACTGG 2259
 : : : : :
 QY 538 Y 538
 Db 2260 A 2260
 : : : : :
 RESULT 13
 AAC77214
 ID AAC77214 standard; cDNA; 2881 BP.
 XX AAC77214;
 XX
 DT 08-FEB-2001 (first entry)
 XX
 DE Human ORFX ORF2769 polynucleotide sequence SEQ ID NO:5537.
 XX
 KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
 KW vulnary; antipsoiatric; antiparkinsonian; nootropic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
 KW antianemic; gene therapy; cancer; proliferative disorder; hyperension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KW thrombosis; contraceptive; ss.
 XX
 OS Homo sapiens.
 XX
 XX
 PN WO200058473-A2.
 XX
 PD 05-OCT-2000.
 XX
 XX 31-MAR-2000; 2000WO-US008621.
 XX
 PF 31-MAR-1999; 99US-0127607P.
 PR 02-APR-1999; 99US-0127636P.
 PR 05-APR-1999; 99US-0127728P.
 PR 30-MAR-2000; 2000US-00540763.
 XX
 XX (CURA-) CURAGEN CORP.
 XX
 XX Shimkets RA, Leach M;
 PI
 XX WPI; 2000-602362/57.
 DR P-PSDB; AAB43005.
 DR
 XX
 PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease.
 XX
 PS Claim 5; Page 4718-4720; 5507pp; English.
 XX
 CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
 CC antipsoiatric; antiparkinsonian; nootropic; neuroprotective; osteopathic;
 CC anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;
 CC cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
 CC dermatological; immunosuppressive; antiinflammatory; antibacterial;
 CC antiviral; antifungal; antirheumatic; antithyroid; and antianemic. The
 CC sequences can be used for determining the presence of or predisposition
 CC to, or preventing or treating pathological conditions associated with an
 CC ORFX-associated disorder. The nucleic acids can be used to express ORFX

XX OS Homo sapiens.
 XX PN WO200153312-A1.
 XX PD 26-JUL-2001.
 XX PF 26-DEC-2000; 2000WO-US034263.
 XX PR 23-DEC-1999; 99US-00471275.
 XX PR 21-JAN-2000; 2000US-00488725.
 XX PR 25-APR-2000; 2000US-00552317.
 XX PR 20-JUN-2000; 2000US-00598042.
 XX PR 19-JUL-2000; 2000US-00620312.
 XX PR 03-AUG-2000; 2000US-00653450.
 XX PR 14-SEP-2000; 2000US-00662191.
 XX PR 19-OCT-2000; 2000US-00693036.
 XX PR 29-NOV-2000; 2000US-00727344.
 XX (HYSE-) HYSEQ INC.
 XX PA Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 XX PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao Q;
 XX PI Zhou P, Goodrich R, Drmanac RT;
 XX XX
 XX DR WPI; 2001-442253/47.
 XX DR P-PSDB; AAM39450.
 XX XX
 XX PT Novel nucleic acids and polypeptides, useful for treating disorders such
 XX PT as central nervous system injuries.
 XX XX
 XX PS Claim 1; SEQ ID NO 809; 10078pp; English.
 XX CC The invention relates to human nucleic acids (AAI57798-AAI61369) and the
 XX CC encoded polypeptides (AAM38642-AAM42213) with nootropic,
 XX CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 XX CC in gene therapy. A composition containing a polypeptide or polynucleotide
 XX CC of the invention may be used to treat diseases of the peripheral nervous
 XX CC system, such as peripheral nervous injuries, peripheral neuropathy and
 XX CC localised neuropathies and central nervous system diseases, such as
 XX CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 XX CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 XX CC utilisation of the activities such as: Immune system suppression,
 XX CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 XX CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 XX CC assays for receptor activity, arthritis and inflammation, leukaemias and
 XX CC C.N.S disorders. Note: The sequence data for this patent did not form
 XX CC part of the printed specification
 XX SQ Sequence 3001 BP; 662 A; 854 C; 846 G; 639 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 6,79e-57 Length: 3001
 Score: 1020.00 Matches: 220
 Percent Similarity: 55.5% Conservatives: 69
 Best Local Similarity: 42.2% Mismatches: 151
 Query Match: 35.7% Indels: 81
 DB: 4 Gaps: 12

US-10-635-977-2 (1-541) x AAI58606 (1-3001)

114 PheSerGlnCysGlnAlaLeuLeuAsnArgIleThrSerValAsnProGlnThrAspIle 133
 1484 GTCCAGCGCTGTGAGCATCTCTGAGCAGCTGCGAGCGCTGTGTACCCAGATAGACATG 1543
 134 AspGlyLeuArgAsnIleTrpIleIleLysProAlaAlaLysSerArgGlyArgAspIle 153
 1544 GAAGGGGATCGCAACATCTGGATCTGAAGCCAGGAGCCAACTCCCGCGAGGAGCATC 1603
 154 ValCysMetAspArgValGluGluLeuLeuAlaAlaAspHisProLeuSer 173
 1604 ATGTGCATGGACACCTGGAGAGATGCTGAAGCTGGTGAACGGCAACCCCTGGTGATG 1663
 174 ArgAspAsnLysTrpValValGlnLysTrpIleGluThrProLeuLeuIleCysAspThr 193
 1664 AAGACCGCAAGTGGTGTGAGAGATATATTGAGCGGCCCTCTCTCATCTTTGGCACC 1723
 194 LysPheAspIleArgGlnTrpPheLeuValThrAspTrpAsnProLeuThrIleTrpPhe 213
 1724 AAGTTTGACTCAGACAGTGTCTCTGGTAACTGACTGGAACCCACTTACCGCTGTGTTTC 1783
 214 TyrLysGluSerTyrLeuArgPheSerThrGlnArgPheSerLeuAspLysLeuAspSer 233
 1784 TACCGCAGCAGCTATATCCGCTTTTCCACGACGCCCTCTCTCCCTGAAGAACCTGGACAAC 1843
 234 AlaIleHisLeuCysAsnAsnAlaValGlnLysTyrLeuLysAsnAspValGlyArgSer 253
 1844 TCAGTGCACCTGTGCACCAACTCCATCCAGAGACCTCGAGAACCTCATGCCATCGGCAT 1903
 254 ProLeuLeuProAlaHisAsnMetTrpThrSerThrArgPheGlnGluTyrLeuGlnArg 273
 1904 CCACCTGCTTCCGCCACACAACTGTGGTGTAGCCAGAGGTTCAGGCGCCACCTCGCAGGAG 1963
 274 GlnGlyArgGlyAlaValTrpGlySerValIleTyrProSerMetLysLysAlaIleAla 293
 1964 ATGGTGCCCAAAATGCTTGGTCCACCATCATCTGCTGGTGGTGAAGGATGCTGTGATC 2023
 294 HisAlaMetLysValAlaGlnAspHisValGluProArgLysAsnSerPheGluLeuTyr 313
 2024 CACGCACTTCAGACCTCCAGACACCGTGCAGTGTGCGAAGCCAGCTTTGAGCTCTAT 2083
 314 GlyAlaAspPheValLeuGlyArgAspPheArgProTrpIleLeuLeuAsnSerSer 333
 2084 GGGCTGACTTCTGTTCGGGGAGGACTTCCAGCCCTGGCTGATTGAGATCAACGGCAGC 2143
 334 ProThrMetHisProSerThrProValThrAlaGlnLeuCysAlaGlnValGlnGluAsp 353
 2144 CCCACATGGCACCCCTCCACAGCAGTCACTGCCCGGCTGTGTGCGTGCAGCTGAC 2203
 354 ThrIleLysValAlaVal-----AspArgSerCysAspIleGlyAsnPheGlu 369
 2204 ACCCTGGCGTGTGTCATTGACCGGATGCTGGACCGCAACTGTGACACAGGCGCTTTGAG 2263
 370 LeuLeuTrpArgGlnProVal---ValGluPro-ProProPheSerGlySerAspLeuCy 388
 2264 CTCATCTATAAGCAGCGCCGTCACACTTCCCGAGCTCCACACCAAGGC----- 2312
 388 sValAlaGlyValSerValArgAlaArgGlnValLeuProValCysAsnLeuLy 408
 2312 ----- 2312
 408 sAlaSerAlaSerLeuLeuAspAlaGlnProLeuLysAlaArgGlyProSerAlaMetPr 428
 2313 -----CCAGTGCCTTCTCCCATGTACTCCGACACCGAGGCCAGGTCTCTCAGACGACAGC 2368
 428 oAspProAlaGlnGlyProProSerProAlaLeuGlnArgAspLeuGlyLeuLysGluGl 448
 2369 ACAGCAAGCTGGTGGGCACTAAGGCCCTGTGCGACCAACAGGCAA---GGCCTTGAGGACTC 2425
 448 uLys-----GlyLeuProLeuAlaLeuLeuAlaPro---LeuArgGlyAla----- 462
 2426 TACCACCGGCTAAGGTCTTCATTTCCCTCCACCGAACCTTGATTTCAAGGTGGCACCCA 2485
 463 -----AlaGluSerGlyGlyAlaAlaGlnProThrArgThr----- 474

Db 2486 GCATCTGAAGCAAGAAAGGTGGCTCGACCTGTGACTCAACCCAGTGGACAGTGCT 2545
 QY 475 -----LysAlaAlaGlyLysValGluLeuProAlaCysProCysAr 488
 Db 2546 GAGCAGGGGTGAGGGCTGAGGGACAGAGGGAGCTCCAGGCTGG----- 2597
 QY 488 gHisValAspSerGlnAlaProAsnThrGlyValProValAlaGlnProAlaLysSerTr 508
 Db 2598 -----CTGGCACCCCAAGGGAAGAGCTG 2620
 QY 508 pAspProAsnGlnLeuAsn-----AlaHisProLeuGluProValLe 522
 Db 2621 GTCTCCTCAGAGGCCCTCTCTCCACAGACTTCTGATCATCTCCTCTCTCCCTCTCT 2680
 QY 522 uArgGlyLeuLysThrAlaGluGlyAlaLeuArgPro-----ProProGlyG1 538
 Db 2681 T-----TCACACCGAGGCTCTGCTCTCTGCTCGGAGGCCCTCCAGCTGG 2728
 QY 538 Y 538
 Db 2729 A 2729
 RESULT 15
 ADQ98824
 ID ADQ98824 standard; cDNA; 3001 BP.
 AC ADQ98824;
 DT 23-SEP-2004 (first entry)
 DE DNA encoding human GPCR-like protein seqid 494.
 KW ophthalmological; immunomodulatory; cytostatic; antiatherosclerotic;
 KW antidiabetic; GPCR-like protein; ophthalmic disorder;
 KW neurological disorder; immunological disorder; nephritic disorder;
 KW hormonal dysfunction; cancer; atherosclerosis; diabetes;
 KW molecular weight marker; food supplement; human; ss.
 OS Homo sapiens.
 PN US6569662-B1.
 PD 27-MAY-2003.
 PF 19-JUL-2000; 2000US-00620312.
 PR 21-JAN-2000; 2000US-00488725.
 PR 25-APR-2000; 2000US-00552317.
 XX (HYSE-) HYSEQ INC.
 XX Tang YT, Zhou P, Drmanac RT;
 XX WPI; 2001-442255/47.
 XX New G-protein-coupled receptor-like polypeptides and polynucleotides,
 PT useful for treating diseases of ophthalmic, neurological, immunological
 PT and nephritic systems and hormonal dysfunction, cancer, atherosclerosis
 PT and diabetes.
 XX Example 2; SEQ ID NO 494; 92pp; English.
 XX The invention describes an isolated polynucleotide (I) comprising a fully
 CC defined (SI) of 749, 3188, 2484, 1169, 2936, 1467, 5773, 5714, 4041,
 CC 1372, 3996, 3945, 1735, 1788, 585, 1782, 927, 5714 or 2282 nucleotides as
 CC given in the specification, its translated or protein coding portion, its
 CC extracellular portion or its active domain. The GPCR-like polypeptides
 CC and polynucleotides are useful for the treatment of diseases of
 CC ophthalmic, neurological, immunological and nephritic systems. They may
 CC also be used to treat hormonal dysfunction, cancer, atherosclerosis and
 CC diabetes. The antibodies are useful for detecting or quantitating the
 CC polypeptide in tissue. The polypeptides can also be used as molecular

CC weight markers and as a food supplement. This sequence represents a human
 CC polynucleotide of the invention.

SQ Sequence 3001 BP; 662 A; 854 C; 846 G; 639 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 6,79e-57 Length: 3001
 Score: 1020.00 Matches: 220
 Percent Similarity: 55.5% Conservative: 69
 Best Local Similarity: 42.2% Mismatches: 151
 Query Match: 35.7% Indels: 81
 DB: 5 Gaps: 12

US-10-635-977-2 (1-541) x ADQ98824 (1-3001)

QY 54 GlnLeuValAspIleAlaCysLysValCysGlnAlaTyrLeuGlyGlnLeuGluHisGlu 73
 Db 1304 GAGTTTGGTGAAGTGAAGCTGTGTGTCGCGAGGAGTACCTTAGCAACTTGGCCCATG 1363
 QY 74 AspIleAspThrSerAlaAspAlaValGluAspLeuThrGluAlaGluTyrGluAspLeu 93
 Db 1364 GACATCGACAGGACCTCGAGGCCCGCTGTACCTCACCCCGAGGGCTGGTCCCTCTTC 1423
 QY 94 ThrGlnGlnTyrSerLeuValHisGlyAspAlaPheIleSerAsnSerArgAsnTyr 113
 Db 1424 CTCCAGCGCTACTCAAGTGTGTCCAGAGGGGCGAAGCTCAGGCACCTCGACACTCAG 1483
 QY 114 PheSerGlnCysGlnAlaLeuAsnArgIleThrSerValAsnProGlnThrAspIle 133
 Db 1484 GTCCAGCGCTGTGAGGACATCTCTGAGCGTGCAGCGGTGTGTTACCCAGATGACATG 1543
 QY 134 AspGlyLeuArgAsnIleTyrIleLysProAlaAlaLysSerArgGlyArgAspIle 153
 Db 1544 GAAGGGATCGACACATCTGATCGTGAAGCAGCAGGACCAAGTCCCGCGGAGGCGATC 1603
 QY 154 ValCysMetAspArgValGluGluLeuLeuGluLeuAlaAlaAspHisProLeuSer 173
 Db 1604 ATGTGCATGAGCACCTCGAGGAGATGTGAAGCTGTGTGAAGCGCAACCCCGTGTGATG 1663
 QY 174 ArgAspAsnLysTyrValValGlnLysTyrIleGluThrProLeuLeuLysCysAspThr 193
 Db 1664 AAGGACGCAAGTGGGTGTGAGAAATATTGAGCGGCCCTCTCTATCTTTGGCAC 1723
 QY 194 LysPheAspIleArgGlnTyrPheLeuValThrAspTyrAsnProLeuThrIleTyrPhe 213
 Db 1724 AAGTTTGACCTCAGACAGTGTCTCTGTACTGACTGGAACCCACTTACCGTGTGGTTC 1783
 QY 214 TyrLysGluSerTyrLeuArgPheSerThrGlnArgPheSerLeuAspLysLeuAspSer 233
 Db 1784 TACCGCAGACGATATATCCGCTTTTCCAGCGAGCCCTTCTCCCTGAAGAACCTGGACAAC 1843
 QY 234 AlaIleHisLeuCysAsnAsnAlaValGlnLysTyrLeuLysAsnAspValGlyArgSer 253
 Db 1844 TCAGTGACCTGTGCAACAACTCCATCCAGAAAGCACTGGAGAACTCATATGCGCATC 1903
 QY 254 ProLeuLeuProAlaHisAsnMetTyrThrSerThrArgPheGlnGluTyrLeuGlnArg 273
 Db 1904 CCATGTCTCCGCCAGACACACATGTGTGTAGCAGAGGTTCAGGCCACCTGCGAGGAG 1963
 QY 274 GlnGlyArgGlyAlaValTyrGlySerValIleTyrProSerMetLysLysAlaIleAla 293
 Db 1964 ATGGGTGCCCAATGTGTGTCACCATCATCTGCTGCTGCTGATGAAGGATGCTGTGATC 2023
 QY 294 HisAlaMetLysValAlaGlnAspHisValGluProArgLysAsnSerPheGluLeuTyr 313
 Db 2024 CACGCATCTCAGACCTCCAGAGACACCGTGCAGGTGTGCGAAGGCCAGCGCTTTGAGCTCAT 2083
 QY 314 GlyAlaAspPheValLeuGlyArgAspPheArgProTyrLeuIleGluLeuAsnSerSer 333
 Db 2084 GGCCTGACTTGTGTGCGGAGGAGACTTCCAGCCCTGGCTGTGATGATCAACCCAGC 2143
 QY 334 ProThrMetHisProSerThrProValThrAlaGlnLeuCysAlaGlnValGlnGluAsp 353

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame plus p2n model

Run on: April 4, 2006, 09:36:08 ; Search time 7900 Seconds

(without alignments)

3892.697 Million cell updates/sec

Title: US-10-635-977-2

Perfect score: 2854

Sequence: 1 MASSILKVVVSHQSCSRSSR.....LRGLKTARGALRPPPGKGKS 541

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_p2n.model -DB=xlh
-Q=/abss/ABSSWEB_spool/US10635977/runat_04042006_103531_9993/app_query.fasta_1
-DB=GenEmbl -QFMT=fastcap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000 -HOST=abss603h
-USER=US10635977 @CGN_1_1_4939 @runat_04042006_103531_9993 -NCPU=6 -ICPU=3
-NO_MMAP -NEG_SCORES=0 -WAIT -DGBLOCK=100 -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:

1: gb_ba.*
2: gb_in.*
3: gb_env.*
4: gb_em.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pr.*
9: gb_ro.*
10: gb_scs.*
11: gb_sy.*
12: gb_un.*
13: gb_vl.*
14: gb_htg.*
15: gb_pl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	1802	63.1	101270	8	HS355C18
2	1455	51.0	817	6	CQ724907
3	1075	37.7	4238	8	HS805098
					AL022327 Human DNA
					CQ724907 Sequence
					AL833939 Homo sapi

4	1070.5	37.5	2380	6	AX301197	Sequence
5	1056.5	37.0	2097	8	BC098298	Homo sapi
6	1054	36.9	2326	6	BD158727	BD158727 Primer fo
7	1054	36.9	2326	6	AX881015	Sequence
8	1054	36.9	2326	8	AX881015	Sequence
9	1020	35.7	2553	8	AK023960	Homo sapi
10	1020	35.7	2553	8	HS800637	Homo sapi
11	1001.5	35.1	3001	6	AR339003	Sequence
12	1000.5	35.1	1769	8	BC099735	Sequence
13	983.5	34.5	1743	8	BC098361	Homo sapi
14	983.5	34.5	2848	6	AX834642	Sequence
15	960.5	33.7	2848	8	AK097236	Homo sapi
16	942	33.0	1832	4	BT021865	Bos tauru
17	942	33.0	205949	14	AC134940	AC134940 Rattus no
18	940.5	33.0	218249	14	AC097425	AC097425 Rattus no
19	940.5	33.0	5282	6	CQ842940	Sequence
20	922	32.3	1684	8	AK125875	Homo sapi
21	880	30.8	1897	9	AF078842	Homo sapi
22	856	30.0	163783	9	BC006830	Mus muscu
23	772	27.0	3828	6	AC119959	Sequence
24	772	27.0	3828	8	CQ850619	Sequence
25	733	25.7	2979	6	AK127786	Homo sapi
26	556.5	19.5	5728	6	CQ595967	Sequence
27	556.5	19.5	5728	2	CQ595967	Sequence
28	556.5	19.5	7854	2	AC003052	Sequence
29	556.5	19.5	86398	14	AC003052	Drosophill
30	556.5	19.5	167201	2	AC092397	AC092397 Drosophill
31	536.5	18.8	2543	6	AE003614	AE003614 Drosophill
32	535.5	18.8	3914	2	CQ595656	Sequence
33	524	18.4	1813	9	BT021443	Drosophill
34	514.5	18.0	4615	6	BC097453	Rattus no
35	514.5	18.0	76854	2	CQ595655	Sequence
36	514.5	18.0	86398	14	AC003052	Drosophill
37	514.5	18.0	167201	2	AC092397	AC092397 Drosophill
38	514.5	18.0	226965	9	AC153589	Mus muscu
39	514.5	18.0	259718	2	AE003614	AE003614 Drosophill
40	501.5	17.6	224077	14	AC120010	AC120010 Mus muscu
41	486	17.0	754	10	BV579068	BV579068 G591P6143
42	480.5	16.8	1958	6	BD160681	BD160681 Primer fo
43	480.5	16.8	1958	8	AX884044	AX884044 Sequence
44	480.5	16.8	1958	8	AK024110	Homo sapi
45	478.5	16.8	145435	14	AC026685	AC026685 Homo sapi

ALIGNMENTS

RESULT 1	HS355C18	101270 bp	DNA	linear	PRI 18-MAY-2005
LOCUS	Human DNA	sequence from clone RP3-355C18 on chromosome 22q13.3			
DEFINITION	Contains the KIAA0027 gene, ESTs, STSs, GSSs and seven putative CpG islands, complete sequence.				
ACCESSION	AL022327				
VERSION	AL022327.17	GI:5304851			
KEYWORDS	HTG; CpG island; KIAA0027.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;				
	Hominidae; Homo.				
REFERENCE	1 (bases 1 to 101270)				
AUTHORS	Cobley, V.				
TITLE	Direct Submission				
JOURNAL	Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton, Cambridgehire, CB10 1SA, UK. E-mail enquiries: vga@sanger.ac.uk				
COMMENT	Clone requests: clonerequest@sanger.ac.uk				
	On Jun 30, 1999 this sequence version replaced gi:5262834.				
	The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:				
	Em., EMBL; Sw., SWISSPROT; Tr., TrEMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at				
	http://www.sanger.ac.uk/Projects/Celegans/wormpep This sequence was generated from part of bacterial clone contigs of human				

Db	46863	TCCGATCAGAGGCAGGATCCGGGTTGATGGGGTTAATCAGGATTTCCCTCCTTTGTATCA	46804
QY	151	-----	151
Db	46803	CAAACTGGTCTATGCTTTCCTTTTAGTTTATTTATTTATTTATTTATTTATTTAGTTTTT	46744
QY	151	-----	151
Db	46743	TGAACGAGTTTCACTCTGTGTTGCCAGGCTGAAGTGCAGTGGCGCAATCTTGCCCTCAC	46684
QY	151	-----	151
Db	46683	TGCAACTCCACCTCCAGGTTTAAGCAATTTCTCTGCCCTTAGCCTCCCAAGTAGTGGG	46624
QY	151	-----	151
Db	46623	ACTACAGGGCCACACACCCGGCTAAATTTTTTGTATTTTATTTAGTAGAGTGGATTTCA	46564
QY	151	-----	151
Db	46563	CCATGTTGCCAGGCTGCTTTGAACCTCCGACCTCAAGTGAATCACCTGCCTCGGCCTC	46504
QY	151	-----	151
Db	46503	CCAAAGTGTGGGATTACAGGCATTAAGCCACCAATGCCATCTCTGGCTCCCTTTTAA	46444
QY	151	-----	151
Db	46443	AGTGCTCTGTCCCTGCCCTTCACTGATCTCTGTGTCCCTCTGGGTCTCAGAGGCTG	46384
QY	151	-----	151
Db	46383	GTTCGTTGGGGAGAGCCTGGGGACTCAGGCACTGGCTTTAGCCTCATCATGAGGC	46324
QY	151	-----	151
Db	46323	TAAGCCTTCCCTGCTCCCATGCCCCACCAACCATGCTGCTGCTGGCCTTGGACAA	46264
QY	151	-----	151
Db	46263	AGCCAGTGCAGTCTCAGGGCCCAAGCTGCGACCTGCCCGGCTCAGGAGCTCAGGTC	46204
QY	152	-----AspIleValCys	155
Db	46203	CCGCTGCACAGTGGCTGTGCATGCTGGGTTGAGCGTGGCTCTTTGCACACATAGTGC	46144
QY	156	MetAspArgValGluGluIleLeuGluLeuAlaAlaAspHisProLeuSerArgAsp	175
Db	46143	ATGGACCGTGTGGAGGAGATCCTGGAGCTGGCAGCTGCAGACCACTCTTTCCAGGGAC	46084
QY	176	AsnLysTyrValValGlnLysTyrIleGluThrProLeuLeuIleCysAspThrLysPhe	195
Db	46083	AACAAGTGGTGGTCCAGAGTACATCGAGACCGCTGCTCATCTGTGACACCAAGTTC	46024
QY	196	AspIleArgGlnTrpPheLeuValThrAspTrpAsnProLeuThrIleTrpPheTyrLys	215
Db	46023	GACATCAGACAGTGGTCTCTGCTCAGGACTGGAACCCCTTGACCATCTGGTTCTACAG	45964
QY	216	GluSerTyrLeuArgPheSerThrGlnArgPheSerLeuAspLysLeuAsp	232
Db	45963	GAGAGTTACTTGGGTTCTCACTCAGGCTTCTCCCTGGACAAGCTGGACAGTGCAGTG	45904
QY	232	-----	232
Db	45903	CGTGGGCGACCTGGACACCTGGGCGCAGGGAAATGGGCTGCTGCTGATGCTGGAGCA	45844
QY	232	-----	232
Db	45843	GCCGAGTGAGAGCAGCAGCGGCGGCCCCCTCACTGCCACCTGCCACCACGAGAGGC	45784
QY	232	-----	232
Db	45783	ACAGATGGGGAACCTGAGGCTTGGGAGGCTGATGATCTCCCCGAGGACAGGCCAGGCC	45724
QY	232	-----	232
Db	45723	CACCGGCTTCAAGCCAGCGCCTACACCCACGCTCTGTGATCATCTCTGGAATAATC	45664
QY	232	-----	232
Db	45663	TTAAAGGCACCATTTTGCTGTGCCAGGGGATTTGAATGTTGTGGGAGCACCCTGGG	45604
QY	232	-----	232
Db	45603	CCCTGGCGCAGGGGAGCGGCTGTGGGTGCTGTGGAGGCTCAGATATCCGCAAGTG	45544
QY	232	-----	232
Db	45543	TGCTGCCAGGCTGGATGGACCCCACTCACCCGGGTCCCGCTGGGAGGAGGCGCA	45484
QY	232	-----	232
Db	45483	GAGCCTGTGGAGTCTCAGGAGTGTCTGTGCACACAGCGAGCCTGCCACCTCTGTCT	45424
QY	232	-----	232
Db	45423	GCACCAAGCTTCTCTGACTATTGTGTGGCCCTTGTGTTGGTGTGAGCGCTAGGGGACC	45364
QY	232	-----	232
Db	45363	TGACGGCTGGGCGAGGAGGGGTCTGCTTTTACGAGTGAGAATCCAGGAGGCTTGAGA	45304
QY	232	-----	232
Db	45303	GTCACGCGACACAGACCTTTCAGGGAGTTAACTCCGGCCAGGCGAGCGCTGCGGTCA	45244
QY	232	-----	232
Db	45243	TGCTGTGGTGGTTCATGCTCGAAGTCACACCCAGAGAGCCCTGGGAGGAGTGGCCAGGG	45184
QY	232	-----	232
Db	45183	GCTACGACCTCACCCTGCTGCTCACTGCTCTGTGTTCAAGTGGGCTTGGAGGGGTCCAGGAAC	45124
QY	232	-----	232
Db	45123	TCACCTGTGCAAAAGCAGCTCCAGGCGAGCAGTGAGAAGCAGCGGCCCTTGCCAGCG	45064
QY	232	-----	232
Db	45063	CCCATTTCTGGGCTTCCAGCTGTGAGCTGTGCGCTTGCCCAATGCCATGGCCTCTGTGAG	45004
QY	232	-----	232
Db	45003	CCCTGTTTCTCATCTGGGACAGTAACGCTCATCCAGCTGGGCGCTCCAGCAGAGGCT	44944
QY	232	-----	232
Db	44943	TGGAGTCTCAGGGCAGGGTTCTCTAGAGCAGCTGGTGGGCTCAGCACGGCTGGGAG	44884
QY	232	-----	232
Db	44883	GACCGGCTTGGGTGCCCTCTCTGGGCCAAGGAGGCTGAGCACCTCGTCCCTCGTCCCCC	44824
QY	233	---SerAlaIleHisLeuCysAsnAlaValGlnLysTyrLeuLysAsnAspValGly	251
Db	44823	TGCAGCGCATCCACTGTGCAACACCCCGTCCAGAGTACCTGAGAATGATGTGGGC	44764
QY	252	ArgSerProLeuLeuProAlaHisAsnMetTrpThrSerThrArgPheGlnGluTyrLeu	271
Db	44763	CGCAGCCCCCTGCTGCCCGCCACACAAATGTGGACCAACAGGTTCCAGGAGTACCTG	44704
QY	272	GlnArgGlnGlyArgGlyAlaValTrpGlySerValIleTyrProSerMetLysAla	291
Db	44703	CAGCGCCAGGGCGCTGGCGCGTGTGGGCGAGCTCATCTACCCGCTCCATGAAGAGGCC	44644

```
QY 292 IleAlaHisAlaMetIysValAlaGlnAspHisValGluProArgIysAsnSerPheGlu 311
Db 44643 ATGCCCCAGCCCATGAAGTGGCCCAAGACCACGTCGGAGCCTCGCAAGAACAGCTTTGAG 44584
QY 312 LeuTyrGlyAlaAspPheValLeuGlyArgAspPheArgProTyrLeuIleGluIleAsn 331
Db 44583 CTCTACGGGGCTGACTTCGTCTTTGGAGAGGACTTCAGGCCCTGGCTGATCGAGATCAAT 44524
QY 332 SerSerProThrMetHisProSerThrProValThrAlaGlnLeuCysAlaGlnValGln 351
Db 44523 TCCAGCCCCACCATGCACCGCTCCACGCGGTCAACGCGCCAGCTGTGTGCACAGGTGCAG 44464
QY 352 GluAspThrIleIysValAlaValAlaAspArgSerCysAspIleGlyAsnPheGluLeu 371
Db 44463 GAGGACACCATCAAGTGGCCGTGGACCGCAGCTGTGACATCGGCAACTTCGAGCTCCTG 44404
QY 372 TTPArg----- 373
Db 44403 TGGAGCAGGTGAGCCACGCCGCCCTCTGGGGACTTTGGTGGCGCCCGCCATGTGGATG 44344
QY 373 ----- 373
Db 44343 CTTTGGCCACCTGCTCACTGGGCGAGGTGGAGCTGCAGGCCGCGCAGTGCAGCGGGTCTA 44284
QY 373 ----- 373
Db 44283 TGGAGGACACCGATCCTGCTCTGTGACCATGTCTCGGTCCCGCTGAGGGGCCCGCCTGCC 44224
QY 373 ----- 373
Db 44223 CTGCTCTCCATGGGGCTCAGTGCCTGCTCGGCACCGAGGAGGGGTACTCCAGGGGTG 44164
QY 373 ----- 373
Db 44163 CGGGCGAGGGCCCGCTCACTCAGCTCAGTGCCTCTGTCCAGCTTTCTCCCTCTTG 44104
QY 373 ----- 373
Db 44103 TTCCACGCTCACCCAGTCTCTGACTGTGTGTAGAGTGGGGGTCCATGCCCGCAGCCTC 44044
QY 373 ----- 373
Db 44043 CAGGCCACCTATGCTGCTCTTGTCCCTTGGCTGCAGTGCCTCCTCCACTCTGAGGATCT 43984
QY 373 ----- 373
Db 43983 TTCGGGGCTCCCTGTGCTGCGGATGTGTGAGTCAATTGGAGCGGTGCCAATGCC 43924
QY 373 ----- 373
Db 43923 TCTCTCTTGGGGCTCTCTGTGGCTAACTCGGGAAAGAGACGTGAGTCCCTGGGTGT 43864
QY 373 ----- 373
Db 43863 GGGGCCCCCGAGGCTGACCCCGAAGCCTCAGGAAGCCTTGTTCCCGCAGGTTCTCCAG 43804
QY 373 ----- 373
Db 43803 AGCTGGGGGCTCTCTGGGGGTCTGCTGGGACCCAGGTGTTTCACTGTTCACTCTGTAG 43744
QY 373 ----- 373
Db 43743 TCAGTGGGGGTCTGTGCCCGCAGCCAGCCCTCTCTGCACACAGCCCTCTCTGCAGGGACCC 43684
QY 374 -----GlnProValValG1 378
Db 43683 CCCCTATCCCGGCCACAGCCAGCACCCCAACCTGTTCTTCCCGACAGCGCGTGTGTGA 43624
QY 378 uProProPheSerGlySerAspLeuCysValAlaGlyValSerValArgAlaAr 398
Db 43623 GCCGCCCATTTACGGGGTCCGACCTCTGCTGGGGGTGTCAGTGTGAGGAGAGCCAG 43564
```

```
QY 398 gArgGlnValLeuProValCysAsnLeuIysAlaSerAlaSerLeuLeuAspAlaGlnPr 418
Db 43563 GAGGCAGGTGCTGCCGCTCTGCAACTCAAGGCTCGGCTCGCTGTGGACGCGCAGCC 43504
QY 418 oLeuIysAlaA:gGlyProSerAlaMetProAspProAlaGlnGlyProProSerProAl 438
Db 43503 GCTGAAGGCACGGGGCCCTCGGCCATGTCAGACCTGCGCCAGGAGCCCCCATCACCAGC 43444
QY 438 alLeuGlnArgAspLeuGlyLeuIysGluIysGlyLeuProLeuAlaLeuLeuAlaPr 458
Db 43443 TCTCCAGCGGGACTTGGGACTGAAGGAAGAGAGGGGCTCCCTCGCTTGTCTGGCACC 43384
QY 458 oLeuArgGlyAlaAlaGluSerGlyGlyAlaAlaGlnProThrArgThrIysAlaAlaG1 478
Db 43383 CTTAAGGGGGGCGAGCGAGCGGTGAGCCGCGCACAGCCACCCGACCAAGCTGCTGG 43324
QY 478 yLysValGluLeuProAlaCysProCysArgHisValAspSerGlnAlaProAsnThrG1 498
Db 43323 GAAGGTGGAGCTCCCGCCCTGTCGCCACGTGAGTGCAGTGCAGGCCCCCAACACCCG 43264
QY 498 yValProValAlaGlnProAlaIysSerTrpAspProAsnGlnLeuAsnAlaHisProLe 518
Db 43263 TGTCCCGGTAGCCAGCCCGCCAAAAGCTGGGATCCAAACACAGCTAAATCGGCACCCGCT 43204
QY 518 uGluProValLeuArgGlyLeuIysThrAlaGluGlyAlaLeuArgProProGlyG1 538
Db 43203 GAGACCTGTGCTGCGGGGCTGAGACAGAGAGGGCGGCTGCGTCCGCCCGCCGAGG 43144
QY 538 yLysGlySer 541
Db 43143 AAAAGGTAAC 43134

RESULT 2
LOCUS CQ724907
DEFINITION Sequence 10841 from Patent WO02068579.
ACCESSION CQ724907
VERSION CQ724907.1 GI:42285764
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Venter,C.J., Adams,M.C., Li,P.W. and Myers,E.W.
TITLE Kits, such as nucleic acid arrays, comprising a majority of humanexons or transcripts, for detecting expression and other uses thereof
JOURNAL Patent: WO 02068579-A 10841 06-SEP-2002;
FEATURES PE Corporation (NY) (US)
source 1. .817
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Alignment Scores:
Pred. No.: 9.97e-68 Length: 817
Score: 1455.00 Matches: 272
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 51.0% Indels: 0
DB: 6 Gaps: 0

US-10-635-977-2 (1-541) x CQ724907 (1-817)

QY 103 GlyAspAlaPheIleSerAsnSerArgAsnTyrPheSerGlnCysGlnAlaLeuLeuAsn 122
Db 2 GGGCATGCTTTCATCTCAATTCCAATTCAAGAAATTAATTTTCGAGTGCAGGCTCTGCTGAAT 61
QY 123 ArgIleThrSerValAsnProGlnThrAspIleAspGlyLeuArgAsnIleTrpIleIle 142
```



```
Db 1687 CAGCAGCTGAGCGCGTGTACCCAGATAGACATGGAAGGGATCGCAACATCTGGATC 1746
Qy 142 IleLysProAlaAlaLysSerArgGlyArgAspIleValCysMetAspArgValGluGlu 161
Db 1747 GTGAAGCCAGGAGCCAAAGTCCCGTGGACGAGGCATCATGTGATGACACCACTTGGAGGAG 1806
Qy 162 IleLeuGluLeuAlaAlaAspHisProLeuSerArgAspAsnLysTrpValValGln 181
Db 1807 ATGCTGAAGCTGTGTGAACGCGCAACCCCGTGTGATGAAGAGACGCGCAAGTGGTGGTCAG 1866
Qy 182 LysTrpIleGluThrProLeuLeuIleCysAspThrLysPheAspIleArgGlnTrpPhe 201
Db 1867 AAGTATATTAGCGCGCCCTCTCACTCTTTGGCACCAGAGTTTGACTCTCAGACAGTGGTTC 1926
Qy 202 LeuValThrAspTrpAsnProLeuThrIleTrpPheTrpLysGluSerTrpLeuArgPhe 221
Db 1927 CTGGAACCTGACGTGAACCCACCTACCGTGTGTCTTACCGCAGCAGCTATATCCGCTTT 1986
Qy 222 SerThrGlnArgPheSerLeuAspLysLeuAspSerAlaIleHisLeuCysAsnAsnAla 241
Db 1987 TCCACGCGAGCCCTTCTCCCTGAAGAACCTGGCAACTCAGTGCACCTGTGCAACAACCTCC 2046
Qy 242 ValGlnLysTrpLeuLysAsnAspValGlyArgSerProLeuLeuProAlaHisAsnMet 261
Db 2047 ATCCAGAGACCTGTGGAACACTCATGCCATCGGCATCCACTGCTTCGGCCAGACACATG 2106
Qy 262 TrpThrSerThrArgPheGlnGluTrpLeuGlnArgGlnGlyArgGlyAlaValTrpGly 281
Db 2107 TGGTCTAGCAGAGGTTCCAGGCCACCTGCAGAGATGGTCCCCAATGCTTGGTCC 2166
Qy 282 SerValIleTrpProSerMetLysLysAlaIleAlaHisAlaMetLysValAlaGlnAsp 301
Db 2167 ACCATCATCTGCTGCGCATGAAGATGCTGTGTATCCACGACCTTCAGACCTCCCAAGGAC 2226
Qy 302 HisValGluProArgLysAsnSerPheGluLeuTrpGlyAlaAspPheValLeuGlyArg 321
Db 2227 ACGTGCAGTGTGGAAGGCCAGCTTTGAGCTCTATGGCGCTGACTTCGTGTTGGGGAG 2286
Qy 322 AspPheArgProTrpLeuIleGluIleAsnSerSerProThrMetHisProSerThrPro 341
Db 2287 GACTTCAGCCCTGGCTGATTGAGATCAACGCCAGCCCAACGATGGCACCTCCACAGCA 2346
Qy 342 ValThrAlaGlnLeuCysAlaGlnValGlnGluAspThrIleLysValAlaVal----- 359
Db 2347 GTCACTCCCGGCTCTGTGCTGGCGTCAAGCTGACACCCCTGCGCGTGCATGACCGG 2406
Qy 360 -----AspArgSerCysAspIleGlyAsnPheGluLeuLeuTrpArgGlnProValVal 377
Db 2407 AGGCTGACCGCAACTGTGACACAGGAGCCTTTGAGCTCATCTATAGCAGCCTGCTGTG 2466
Qy 378 GluProProPheSerGlySerAspLeuCysValAlaGlyValSerValArgArgAla 397
Db 2467 GAGTGCTCAATATATGTTGGGCATCCGCTCTCTGTAGAGGGCTTCACCATCAAGAACCC 2526
Qy 398 Arg-----ArgGlnValLeuProValCysAsnLeuLys 408
Db 2527 ATGCGCATGTGTCATCGCGGATGGGGGTCCGCCAGCAGTCCCTCTGTCGACCCACCGA 2586
Qy 409 AlaSerAlaSerLeu-LeuAspAlaGlnProLeuLysAla-ArgGlyProSerAlaMet- 427
Db 2587 GGCTCTGGGAGGCCGTGACCACTTCCCGAGCCTCCACACCAAGGCCAGCTGCTTCT 2646
Qy 428 -----ProAspProAlaGlnGlyProProSerProAlaLeuGln----- 440
Db 2647 CCCCATGTACTCCGACACCGGCGCAGTCTCTCAGACGACGACGACGACGACGAGCTGGTGGC 2706
Qy 441 -----ArgAspLeuGlyLeuLysGluLys-----GlyLeuP 452
Db 2707 ACTAAGGCCCTGTGCACACAGGCAAGCCCTTGAGGACTCTACCCACGGCTTAAGCTTTC 2766
Qy 452 roLeuAlaLeuLeuAlaPro--LeuArgGlyAla-----AlaGluSerG 466
Db 2767 ATTTCCCTCCACCGAACCTTGATTTCAGGTGGCACCACGATCTCTGAAGCCGAAGAAG 2826
```

```
Qy 466 lYgLYAlaAlaGlnProThrArgThr-----L 475
Db 2827 GTGGGCTCGACTGTGACTCATCCAGTGGACAGTGTGACACGGGTGAGGCTGG 2886
Qy 475 ySAlaAlaGlyLysValGluLeuProAlaCysProCysArgHisValAspSerGlnAlaP 495
Db 2887 AGGCACAGCAGAGGGCAGCTCCACGGCTGG----- 2918
Qy 495 roAsnThrGlyValProValAlaGlnProAlaLysSerTrpAspProAsnGlnLeuAsn- 514
Db 2919 -----CTGGCACCACCAAGGAGAGCTGGTCTCCTCAGAAGCCCTT 2961
Qy 515 -----AlaHisProLeuGluProValLeuArgGlyLeuLysThrAlaG 529
Db 2962 CTTCCACAGACTTCTGATCATCTCCTCTTCTCCCTCTCT-----TCACCC 3009
Qy 529 luGlyAlaLeuArgPro-----ProProGlyGly 538
Db 3010 GAGGCTCTGCTCTCTGCTCCCGAGGCCCCAGCTGGA 3050

RESULT 4
AX301197 2380 bp DNA linear PAT 30-NOV-2001
LOCUS Sequence 39 from Patent WO0185942.
ACCESSION AX301197
VERSION AX301197.1 GI:17382228
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
REFERENCE
1 Yue,H., Tang,Y.T., Au-Young,J., Lu,D.A., Baughn,M.R., Hillman,J.L.,
Azimzai,Y., Lal,P., Yao,M.G., Bandman,O., Burford,N., Batra,S.,
Kearney,L. and Pollicky,J.L.
Cytoskeleton-associated proteins
Patent: WO 0185942-A 39 15-NOV-2001;
Incyte Genomics, Inc. (US)
FEATURES
Location/Qualifiers
1..2380
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
/notes="Incyte ID No: 2156553CB1"
ORIGIN
Alignment Scores: 4.05e-47 Length: 2380
Pred. No.: 1070.50 Matches: 250
Score: 54.8% Conservative: 81
Percent Similarity: 41.4% Mismatches: 198
Best Local Similarity: 37.5% Indels: 76
Query Match: 6 Gaps: 12
DB:
US-10-635-977-2 (1-541) x AX301197 (1-2380)
Qy 2 AlaSerSerIleLeuLysTrpValValSerHisGlnSerCysSerArgSerArgSer 21
Db 285 GCCCGCAACGTTCTCAAGCTGGTGGTGAAGTCTGAG-----TGAAGTCA 329
Qy 22 LysProArgAspGlnArgGluAlaGlySerSerAspLeuSerArgGlnAspAla 41
Db 330 TACCCTATTTCAGGCGAGTAGAGGAGGCGCTTCAGAGCAAGCAGCAGCCAAAGAACAGAG 389
Qy 42 GluAsnAlaGluAlaLysLeuArgGlyLeuProGlyGlnLeuValAspIleAlaCysLys 61
Db 390 AAAAACCCAGTGTG-----GTGTCCCGAGAGTTGTGGATGAAGCTCTGTGT 437
Qy 62 ValCysGlnAlaTrpLeuGlyGlnLeuGluHisGluAspIleAspThrSerAlaAspAla 81
Db 438 GCGTGGCGAGGAGTACCTTAGCAACTTGGCCCACTGACATCGACATCGACAAGGACCTGGAGCC 497
```

QY	82	ValGluAspLeuThrGluAlaGluTrpGluAspLeuThrGlnGlnTyrTyrSerLeuVal	101	QY	421	aArgGlyProSerAlaMetProAspProAlaGlnGlyProProSerProAlaLeuGln	440			
DB	498	CGCTGTACTACCCCGAGGCTGGTCCCTCTTCTCCAGCGCTACTACCAAGTGGTC	557	DB	1578	ACTGGGGCCAGGAGCTGGGGGCACAGTGAAGAGCCAGTCTCCACTGCCACCACTTCAGCC	1637			
QY	102	HisGlyAspAlaPheIleSerAsnSerArgAsnTyrPheSerGlnCysGlnAlaLeuLeu	121	QY	441	-ArgAspLeuGlyLeuGlyGluGlyGlyLeuProLeuAlaLeuLeuAlaProLeuLeu	460			
DB	558	CACGAAGGGCGAAGCTCAGGACCTCCACACTCAGGTCCAGCGCTGTGAGACATCTGTG	617	DB	1638	CCCGAAAGGGGAAGAAAGGCAAGGCGCAAAAGGGCCACAGCCCTGGTCTGCCCAATCTC	1697			
QY	122	AsnArgIleThrSerValAsnProGlnThrAspIleAspGlyLeuArgAsnIleTrpIle	141	QY	460	gGlyAlaAla-----GluSerGlyGlyAlaAlaGlnProThrArgThrLysAlaAl	477			
DB	618	CAGCAGCTGCAGCGGTGTACCCAGATAGACATAGGAAGGGATCGCAACATCTGGATC	677	DB	1698	TGGGAGTGGGATGCCCCAGCAGCAGGATGGGCTGCATTTTC-ACCATGACCTTTCTAG	1756			
QY	142	IleLysProAlaAlaLysSerArgGlyArgAspIleValCysMetAspArgValGluGlu	161	QY	477	aGlyLysValGlu-----LeuProAlaCysProCysArgHisValAspSerGlnAl	494			
DB	678	GTGAAGCCAGGAGCCAAAGTCCGTGGACGAGGCATCATGTGATGACCACTGGAGGAG	737	DB	1757	TGGGACAGGCAACCCCACTTGAACAGATTTGCCACTGAGTCCGAAGAACCCCAAGC	1816			
QY	162	IleLeuGluLeuAlaAlaAspHisProLeuSerArgAspAsnLysTrpValValGln	181	QY	494	aProAsnThrGlyValProValAlaGln-----ProAlaLys	506			
DB	738	ATGCTGAAGCTGTGAACGGCAACCCCGTGGTGAAGGACGGCAAGTGGGTGGTCAG	797	DB	1817	CTGGGTAGACCAATTCCTCCCAAAACACCCGAGTGTTCAGGCGGATTTATTCCTGCTCT	1876			
QY	182	LysTrpIleGluThrProLeuLeuIleCysAspThrLysPheAspIleArgGlnTrpPhe	201	QY	506	sSerTrpAspProAsnGlnLeuAsnAlaHisPro-----	517			
DB	798	AAGTATATTGAGCGGCCCTCTCTATCTTTGGCACCAGTTTGACCTCAGACAGTGGTTC	857	DB	1877	CCAGGCCCTCCCAACCACTGGATCAGCCACCCCAAGAGCCACAGTAGCAAGTA	1936			
QY	202	LeuValThrAspTrpAsnProLeuThrIleTrpPheTyrLysGluSerTyrLeuArgPhe	221	QY	517	-----	517			
DB	858	CTGGTAACCTGACTGGAAACCACTTACCGTGTGGTCTACCGGCAGACGTATATCCGCTTT	917	DB	1937	AAAGCCACTACTCAAAAGTATTGTTTAAATAATACACAGCCAAATAGCTGGGACGGTG	1996			
QY	222	SerThrGlnArgPheSerLeuAspLysLeuAspSerAlaIleHisLeuCysAsnAla	241	QY	518	-LeuGluProValLeuArgGlyLeuLysThrAlaGlyAlaLeuArgProProGln	537			
DB	918	TCCAGCGACCCCTTCTCCCTGAAGAACCTGGCAAACTCAGTCACCTGTGCAACAACTCC	977	DB	1997	GTGTGAGCTGTGGTCCAGGCTACTCGCGAGGCTAATGAGGAT---CGCTTGAGCCCGG	2053			
QY	242	ValGlnLysTrpLeuLysAsnAspValGlyArgSerProLeuLeuProAlaHisAsnMet	261	QY	537	YGLYLYSGLY	540			
DB	978	ATCCAGAGCCTTGGAAGACTCATGCCATCGCATCCACTCTTCGCCGACACACATG	1037	DB	2054	AGGTCAAGGC	2063			
QY	262	TrpThrSerThrArgPheGlnGluTyrLeuGlnArgGlnGlyAlaValTrpGly	281	RESULT 5						
DB	1038	TGGTCTAGCCAGAGTTCAGGCCCACTCGCAGGAGATGGGTGCCCAAACTCTGGTCC	1097	LOCUS	BC098298					
QY	282	SerValIleTyrProSerMetLysLysAlaIleAlaHisAlaMetLysValAlaGlnAsp	301	DEFINITION	BC098298	2097 bp	mRNA			
DB	1098	ACCATCATCGTCCCTGGCATGAAGATGCTGTGATCCAGCATCTCAGACCTCCAGGAC	1157	ACCESSION	BC098298	Homo sapiens tubulin tyrosine	linear			
QY	302	HisValGluProArgLysAsnSerPheGluLeuTyrGlyAlaAspPheValLeuGlyArg	321	VERSION	BC098298.1	transcript variant 1, mRNA	(cDNA clone MGC:120531 IMAGE:40025661),			
DB	1158	ACCGTGCAGTTCGGAAGCCAGCTTGAGCTCTATGGCGCTGACTTGTGTTCGGGAG	1217	KEYWORDS	MGC.	complete cds.				
QY	322	AspPheArgProTrpLeuIleGluIleAsnSerSerProThrMetHisProSerThrPro	341	SOURCE	Homo sapiens (human)					
DB	1218	GACTTCCAGCCCTGGCTGATTGAGATCAAGGCCAGCCACGATGGCACCTCCACAGCA	1277	ORGANISM	Homo sapiens					
QY	342	ValThrAlaGlnLeuCysAlaGlnValGlnGluAspThrIleLysValAlaVal	359	REFERENCE	1 (bases 1 to 2097)					
DB	1278	GTCACTGCCCGCTCTGTGTCGCTGCAAGCTGACACCCCTCGCGGTGCTCATTCACCGG	1337	AUTHORS	Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, P., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Brange, C., Rana, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullaly, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S., Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smal, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.	1397	CONSTRM			
QY	360	-----AspArgSerCysAspIleGlyAsnPheGluLeuLeuTrpArgGlnProValVal	377	TITLE	Mammalian Gene Collection Program Team					
DB	1338	ATGCTGACCCCACTGTGACACAGGAGCCCTTGAGCTCACTATAGCAGCCTCTGTGTG	1397	JOURNAL	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences					
QY	378	GluProProProPheSerGlySerAspLysValAlaGlyValSerValArgArg---	396		Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)					
DB	1398	GAGGTGCTCAATATGTGGGCATCCGCTCTCTGTAGAGGCTTCACCATCAGAGAGCC	1457							
QY	397	-----AlaArgArgGlnValLeuProValCysAsnLeu-----	407							
DB	1458	ATGGCGATGTGTATCGCGGATGGGGGTCCGCCAGCAGTCCCTCTGCTGCCACCGAGA	1517							
QY	408	-----LysAlaSer-AlaSerLeuLeuAspAlaGlnProLeu-----LysAl	421							
DB	1518	GGCTCTGGGGAAGGCAAGGACTCGGGGATCCCTACCCACAGGTCTAGGAAGGC	1577							


```

Db 1383 GCAGTCCCTCTGCTGACCCAGCGAGGCTCTGGGAAGCCGTCACCACTTCCCAGCCTC 1442
Qy 420 sala-ArgGlyProSerAlaMet-----ProAspProAlaGlnGlyProProS 436
Db 1443 CACACCAAGGCCAGCTGCTCTCCCATGTAAGTCTCCGACACACAGGCGCAGGCTCTCAGA 1502
Qy 436 exProAlaLeuGln-----ArgAspLeuGlyLeuLysG 447
Db 1503 CGACAGCAGACCAAGCTGTGGGCACATAAGGCCCTGTGACACACAGGCAAGGCTTGGAG 1562
Qy 447 luGluLys-----GlyLeuProLeuAlaLeuAlaPro---LeuArgGlyAla- 462
Db 1563 ACTCTACCCACGCGTAAGTCTTCAATTCCTCCACCCAGGAGCTTGAATTCAGGTGGCA 1622
Qy 463 -----AlaGluSerGlyGlyAlaAlaGlnProThrArgThr----- 474
Db 1623 CCCAGCATCTCGAAGCCAGAAAGGTGGGCTCGACCTGTGACTCACACCCAGTGGACAG 1682
Qy 475 -----LysAlaAlaGlyLysValGluLeuProAlaCysProC 487
Db 1683 TCGTAGCAGCGGGTCTAGGCTGGAGGCGACAGGAGGCGAGCTCCAGGCTGG- 1738
Qy 487 yeArgHisValAspSerGlnAlaProAsnThrGlyValProValAlaGlnProAlaLys 507
Db 1739 -----CTGGACCCCAAGGGAAGA 1757
Qy 507 exTrpAspProAsnGlnLeuAsn-----AlaHisProLeuGluProV 521
Db 1758 GCTGTCTCTCCCTCAGAAAGCCCTTCTCCACAGACTTCTGATCAATCTCTCTCTCC 1817
Qy 521 aLeuArgGlyLeuLysThrAlaGluGlyAlaLeuArgPro-----ProProG 537
Db 1818 TCCTT-----TCACACGAGGCTCTGCTCTCTGCTGCTCGAGGCCGCCAG 1865
Qy 537 lyGly 538
Db 1866 CTGGA 1870

RESULT 6
BD158727
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,
Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
Primer for synthesizing full-length cDNA and use thereof.
Patent: JP 202191363-A 13570 09-JUL-2002;
HELIX RESEARCH INSTITUTE
OS Homo sapiens (human)
PN JP 2002191363-A/13570
PD 09-JUL-2002
PF 28-JUL-2000 JP 2000280990
PI TOSHIO OTA,TAKAO ISOGAI,TETSUO NISHIKAWA,KOJI HAYASHI,KAORU
PI JUNICHI YAMAMOTO,SHIZUKO ISHII,TOMOYASU SUGIYAMA,AI WAKAMATSU,
PI KEIICHI NAGAI,TETSUJI OTSUKI
PC
C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/PC
10,
PC C12P21/02,C12Q1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
Primer for synthesizing full-length cDNA and use thereof FH Key
Location/Qualifiers
(94)...(2325).
FT CDS Location/Qualifiers
1..2326

```

```

/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

```

ORIGIN

Alignment Scores:

```

Pred. No.: 2,88e-46 Length: 2326
Score: 1054.00 Matches: 233
Percent Similarity: 55.4% Conservative: 79
Best Local Similarity: 41.4% Mismatches: 166
Query Match: 36.9% Indels: 86
DB: 6 Gaps: 12

```

US-10-635-977-2 (1-541) x BD158727 (1-2326)

```

Qy 2 AlaSerSerIleLeuLysTrpValValSerHisGlnSerCysSerArgSerArgSer 21
Db 544 GCCCCCAAGCTTCTCAAGCTGTGTGTAAGTCTGAG-----TGGAAAGTCA 588
Qy 22 LysProArgAspGlnArgGluAlaGlySerSerAspLeuSerSerArgGlnAspAla 41
Db 589 TACCCCTATTTCAGGCGATAGAGGAGAGGCTCAGGAGACACAGCAGCCCAACAGGAG 648
Qy 42 GluAsnAlaGluAlaLysLeuArgGlyLeuProGlyGlnLeuValAspIleAlaCysLys 61
Db 649 AAAAAACCCAGTGTTC-----GTGTCCCGAGAGTTGTGATGAAGCTCTGTGT 696
Qy 62 ValCysGlnAlaTyLeuGlyGlnLeuGluHisGluAspIleAspThrSerAlaAspAla 81
Db 697 GCGTCGAGGAGTACCTTAGCACTTGGCCCACTGGACATCGACAGACCTGGAGGCC 756
Qy 82 ValGluAspLeuThrGluAlaGluTrpGluAspLeuThrGlnGlnTyrTyrSerLeuVal 101
Db 757 CGCTGTACCTCACCCCGAGGGCTGTCCTCTTCTCCAGCGCTACTTACCAAGTGGTC 816
Qy 102 HisGlyAspAlaPheIleSerAsnSerArganTyrPheSerGlnCysGlnAlaLeu 121
Db 817 CACGAAGGGCGAGAACTCAGGCACCTCGACACTCAGGTCCAGCGCTGTGAGACATCTCTG 876
Qy 122 AsnArgIleThrSerValAsnProGlnThrAspIleAspGlyLeuArgAsnIleTrpIle 141
Db 877 CAGCAGCTGCGAGCGCTGTATCCCGATAGATGGAAGGGGATCGCAACATCTGGATC 936
Qy 142 IleLysProAlaAlaLysSerArgGlyArgAspIleValCysMetAspArgValGluGlu 161
Db 937 GTGAAGCCAGGAGCCCAAGTCCCGGAGCAGGSCATCATGTGCATGACACCTGGAGGAG 996
Qy 162 IleLeuGluLeuAlaAlaAspHisProLeuSerArgAspAsnLysTrpValValGln 181
Db 997 ATGCTGAAGCTGTGTAACCGCAACCCCGTGTGTGATGAAGGACGGCAAGTGGTGGTGCAG 1056
Qy 182 LysTrpIleGluThrProLeuLeuIleCysAspThrLysPheAspIleArgGlnTrpPhe 201
Db 1057 AAGTATATTGAGCGGCCCTCTCATCTTGGCCAAAGTTGACCTCAGACAGTGGTTC 1116
Qy 202 LeuValThrAspTrpAsnProLeuThrIleTrpPheTyrLysGluSerTyrLeuArgPhe 221
Db 1117 CTGTGAACCTGACTGGAACCCACTTACCGTGTGTTCTACCGCAGCAGCTATATCCGCTTT 1176
Qy 222 SerThrGlnArgPheSerLeuAspLysLeuAspSerAlaIleHisLeuCysAsnAla 241
Db 1177 TCCAGCGACGCCCTTCTCCCTGAAGAACCTCGACAACTCAGTGCACCTGTGCAACACTCC 1236
Qy 242 ValGlnLysTyrLeuLysAsnAspValGlyArgSerProLeuLeuProAlaHisAsnMet 261
Db 1237 ATCCAGAAAGCACTTGGAGAACTCATGCCATCCGATCCGATCTGCTTCCGCCAGAACATG 1296
Qy 262 TrpThrSerThrArgPheGlnGlyTrpLeuGlnArgGlnGlyArgGlyAlaValTrpGly 281
Db 1297 TGGTCTAGCCAGAGGTTCCAGGCCACCTCGCAGGAGATGGGTGCCCAAAATGCTGGTCC 1356
Qy 282 SerValIleTyrProSerMetLysLysAlaIleAlaHisAlaMetLysValAlaGlnAsp 301

```

	TITLE	Primers for synthesising full-length cDNA and their use
	JOURNAL	Patent: EP 1074617-A 1992-07-FEB-2001;
	RESEARCH	Research Association for Biotechnology (JP)
	FEATURES	Location/Qualifiers
	source	1..2326
		/organism="Homo sapiens"
		/mol_type="unassigned DNA"
	CDS	/db_xref="taxon:9606"
		94..>2325
		/note="unnamed protein product"
		/codon_start=1
		/protein_id="CAE92353.1"
		/db_xref="GI:40035752"
		/translation="MRLNNAKIYVERAVKQKIFITGGCPVIRCLRRRWRVEVKOM VHRGPTLLPPKQLDSSMGSDSTTDEDEDEEPQSQFDFDLFLFKDDLDLGH ALMYGLCLNLRLNLPWFEDVDANSFPFCYCLGAEDDKAFIEDFWLTAAANVLKL VVK SEWSPYFLOAEBSAGDKPKQEKNPVLVSFEVDALCAEEYLNLNHWIDIKO LEAPLYTPGESLSFLQRYOVHEGAEHLRDITQVORCEDILQOLQAUVPOIDMEGD RNIVTWFGAKSRGRCINCMHLEMLKLVNGNVNMKGDWVKKVQIETPLIFGT FDLRQGVTLTDWNPFLTVMFYRDSYIRFSTQPSFLKNJDNVHLNNISIQHLENCHSR HPLPDPNOAHLOSQFOAHLQEMGAPNAWSIIIVPMKDVAVIHALQTSDTVQCRKASF ELYGADFVEDGFOPWLIEINASPTMAPSTAVTARLCAGVOADTLRVVIDRMILDNRCD TGABELIYKPAPSVYPQVVGIRLVEGFTIKKPNAMCHRRMGVRPAVPLTLTORGSGBA RHFFSLHTKAQLFSPHVRHQGVLRQRSHKSLVTGKTALSTTGKALTFTAKVFISL PPNLDKFVAPSILKPRKAPALLCGRPOLEVPCCPLKSEQFIAPVGRSRKPNKSPRP DCDKPRABECMPKRLSLPLKPLPLVLTGTQRRRGDMKLGKPLLRFPFTVALVDLPNNK KQVK"
	ORIGIN	
	Alignment Scores:	
	Pred. No.:	2,88e-46 Length: 2326
	Score:	1054.00 Matches: 233
	Percent Similarity:	55.4% Conservative: 79
	Best Local Similarity:	41.4% Mismatches: 166
	Query Match:	36.9% Indels: 86
	DB:	Gaps: 12
		US-10-635-977-2 (1-541) x AX881015 (1-2326)
Db	Qy	2 AlaSerSerIleLeuLysTrpValValSerHisGlnSerCysSerArgSerSerArgSer 21
Qy	Db	544 GCGCCACAGCTTCCTCAAGCTGTGTGAAGTCTGAG-----TGGAGATCA 588
Qy	Db	22 LysProArgAspGlnArgGluAlaGlySerSerAspLeuSerSerArgGlnAspAla 41
Db	Db	589 TACCCTATTACGCAGTAGAGGAGAGGCCCTCAGGAGACAGCACGCCCCAACAGCAG 648
Qy	Db	42 GluAsnAlaGluAlaLysLeuArgGlyLeuProGlyGlnLeuValAspIleAlaCysLys 61
Db	Db	649 AAAAAACCCAGTCTTG-----GTGTCCCAGAGTTGTGTGGATGAAGCTCTGTGT 696
Qy	Db	62 ValCysGlnAlaTyrlieuGlyGlnleuGluHisGluAspIleAspThrSerAlaAspAla 81
Db	Db	697 GCGTGGAGGAGTACTCTTAGCAACTTGGCCCATATGGACATCGACAAGACCTGAGGCC 756
Qy	Db	82 ValGluAspLeuThrdGluAlaGluTrpGluAspLeuThrGlnGlnTyrtzSerLeuVal 101
Db	Db	757 CCCTGTACTCTACCCCGAGGGCTGTCTCTCTCTCCAGCGCTACTACCAAGTGTCT 816
Qy	Db	102 HisGlyAspAlaPheIleSerAsnSerArgAntyrPheSerGlnCysGlnAlaLeuLeu 121
Db	Db	817 CACGAGGGGCGAGAACCTCAGGCACCTCGACACTCAGGTCCAGCGCTGTGAGGACATCTCT 876
Qy	Db	122 AsnArgIleThrSerValasnProGlnThrAspIleAspGlyLeuArgAsnIleTrpIle 141
Db	Db	877 CAGCAGCTCGAGCGCGTGTGTACCCAGATAGACATGGAAAGGGGATCGCAACATCTGGATC 936
Qy	Db	142 IleLysProAlaAlaLysSerArgGlyArgAspIleValCysMetAspArgValGluGlu 161
Db	Db	937 GTGAACCCAGGAGCCAAAGTCCCGCGGACGAGGCATCATGTGATCGACACCACTGGAGG 996
Qy	Db	162 IleLeuGluLeuAlaAlaAspHisProLeuSerArgAspAsnLysTrpValValGln 181

Db 997 ATGCTGAAGCTGGTGAACGCAACCCGCTGGTGTGATGAAGGACGGCAAGTGGGTGGTGACG 1056
Qy 182 LysTyrlleGluThrProLeuLeuIleCysAspThrLysPheAspIleArgGlnTrpPhe 201
Db 1057 AAGTATATTGAGCGGCCCTCTCTCATCTTTGGCACCAGTTTGACCTCAGACAGTGGTTTC 1116
Qy 202 LeuValThrAspTrpAsnProLeuThrIleTTPheTyrlsGluSerTyrlsLeuArgPhe 221
Db 1117 CTGGTAACTGACTGGAAACCACTTACCGTGTGGTCTTACCGGACAGCTATATCCGCTTT 1176
Qy 222 SerThrGlnArgPheSerLeuAspLysLeuAspSerAlaIleHisLeuCysAsnAla 241
Db 1177 TCCAGCGACCCCTCTCTCCCTGAAGAACCTGGACAACTCAGTCAGCTGTGCACAACTCC 1236
Qy 242 ValGlnLysTyrlsLeuLysAsnValGlyArgSerProLeuLeuProAlaHisAsnMet 261
Db 1237 ATCCGAAGACCACTGGGAACCACTCATGCCATCGGCATCCACTGCTTCGCGCAGACAATG 1296
Qy 262 TrpThrSerThrArgPheGlnTyrlsLeuGlnArgGlnGlyArgGlyAlaValTrpGly 281
Db 1297 TGTCTAGCCAGAGGTTCCAGGCCCACTGCGAGGAGATGGGTGCCCAATGCTTGGTCC 1356
Qy 282 SerValIleTyProSerMetLysLysAlaIleAlaHisAlaMetLysValAlaGlnAsp 301
Db 1357 ACCATCATGCTGCTGGCATGAGGATGCTGTATCCAGCACTTCAGACCTCCAGGAC 1416
Qy 302 HisValGluProArgLysAsnSerPheGluLeuTyrlsGlyAlaAspPheValLeuGlyArg 321
Db 1417 ACCGTGCACTGCGGAAGCCAGCTTTGAGCTCTATGGCGCTGACTTGTGTTCGGGAG 1476
Qy 322 AspPheArgProTrpIleGluLeuIleAsnSerSerProThrMetHisProSerThrPro 341
Db 1477 GACTTCCAGCCCTGCTGANTGAGATCAACGCCAGCCCAACATGGCACTCCACAGCA 1536
Qy 342 ValThrAlaGlnLeuCysAlaGlnValGlnGluAspThrIleLysValAlaVal 359
Db 1537 GTCACTGCCGCGCTGTGTGCTGGTGCAGCTGACACCTCGCGTGTCTATTCACCGG 1596
Qy 360 -----AspArgSerCysAspIleGlyAsnPheGluLeuLeuTrpArgGlnProValVal 377
Db 1597 ATGCTGGACCGCACTGTGACAGGAGCCTTTGAGCTCATATTAAGCAGCCTCTCTGTG 1656
Qy 378 GluProProProPheSerGlySerAspLysCysValAlaGlyValSerValArgAla 397
Db 1657 GAGTGGCTCAATATGTGGGCATCGCGTCTCTGGTAGAGGCTTCACCATCAAGAAG--- 1713
Qy 398 ArgArgGlnValLeuProValCysAsnLeuLysAlaSerAlaSerLeuLeuAspAlaGln 417
Db 1713 ----- 1713
Qy 418 ProLeuLysAlaArgGlyProSerAlaMetProAspProAlaGlnGlyProProSerPro 437
Db 1714 -----CCATGGCGATG----- 1725
Qy 438 AlaLeuGlnArgAspLeuGlyLeuLysGluLysGlyLeuProLeuAlaLeuAla 457
Db 1726 ---TGTCTATCGCGATGGGTGCGC-----CCAGCAGTCCCTCTGCTGT 1767
Qy 458 ProLeuArgGlyAlaAlaGluSerGlyGlyAlaAlaGlnProThrArgTrpLysAlaAla 477
Db 1768 ACCGAGCGAGGCTCGGGAGAGCCGCTCACCCTCCCGAGCTCCACCAAGGCC--- 1824
Qy 478 GlyLysValGluLeuProAlaCysProCysArgHisValAspSerGlnAlaProAsnThr 497
Db 1825 -----CAGTGGCTTCTCC---CCATGTACTCGACACAGGCGGCGCTCTCAGACG 1874
Qy 498 GlyValProValAlaGlnProAlaLysSerTrp-----AspProAsnGln--- 512
Db 1875 -----ACAGCACAGCAAGCTGGTGGGCACTAAGGCCCTGTGCAACACAGGCAAGGC 1925
Qy 513 -----LeuAsnAlaHisProLeuGluProValLeuArg 523
Db 1926 CTTGAGGACTACCCAGGCTAAGGCTTTCATTTCCCTCCACCGCAACCTTGATTTCAA 1985

Qy 524 Gly-----LeuLysThrAlaGluGlyAlaLeuArgPro-----Pro 535
Db 1986 GGTGGCACCAGCATCTCTGAAGCAAGAAAGCTCTCTGCTCTCTGCTCGGAGGCC 2045
Qy 536 ProGlyGly 538
Db 2046 CCAGCTGGA 2054
RESULT 8
AK023960
LOCUS AK023960
DEFINITION Homo sapiens cDNA FLJ13898 fis, clone THYR01001738, weakly similar to TUBULIN--TYROSINE LIGASE (EC 6.3.2.25).
ACCESSION AK023960
VERSION AK023960.1 GI:10436083
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1
AUTHORS Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R., Wakamatsu, A., Hayashi, K., Sato, H., Nagai, K., Kimura, K., Makita, H., Sekine, M., Oobayashi, M., Nishi, T., Shibahara, T., Tanaka, T., Ishii, S., Yamamoto, J., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y., Nagahara, K., Murakami, K., Yasuda, T., Iwayanagi, T., Wagatsuma, M., Shiratori, A., Sudo, H., Hosoiri, T., Kaku, Y., Kodaira, H., Kondo, H., Sugawara, M., Takahashi, M., Kanda, K., Yokoi, T., Furuya, T., Kikkawa, E., Omura, Y., Abe, K., Kamihara, K., Katsuta, N., Sato, K., Tanikawa, M., Yamazaki, M., Ninomiya, K., Ishibashi, T., Yamashita, H., Murakawa, K., Fujimori, K., Tanai, H., Kimata, M., Watanabe, M., Hiraoka, S., Chiba, Y., Ishida, S., Ono, Y., Takiguchi, S., Watanabe, S., Yosida, M., Hotuta, T., Kusano, J., Kanehori, K., Takahashi-Fujii, A., Hara, H., Tanase, I., Nomura, Y., Togiya, S., Komai, F., Hara, R., Takeuchi, K., Arita, M., Imose, N., Musashino, K., Yuki, H., Oshima, A., Sasaki, N., Aotsuka, S., Yoshioka, Y., Matsunawa, H., Ichihara, T., Shiohata, N., Sano, S., Moriya, S., Momiyama, H., Satoh, N., Takami, S., Terashima, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mizoguchi, H., Goto, Y., Shinizu, F., Wakebe, H., Hishigaki, H., Watanabe, T., Sugiyama, A., Takemoto, M., Kawakami, B., Yamazaki, M., Watanabe, K., Kumagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M., Tashiro, H., Tanigami, A., Fujiwara, T., Ono, T., Yamada, K., Fujii, Y., Ozaki, K., Hirao, M., Ohmori, Y., Kawabata, A., Hikiji, T., Kobatake, N., Inagaki, H., Ikema, Y., Okamoto, S., Okitani, R., Kawakami, T., Noguchi, S., Itoh, T., Shigeta, K., Senba, T., Matsumura, K., Nakajima, Y., Mizuno, T., Morinaga, M., Sasaki, M., Togashi, T., Oyama, M., Hata, H., Watanabe, M., Komatsu, T., Mizushima-Sugano, J., Satoh, T., Shirai, Y., Takahashi, Y., Nakagawa, K., Okumura, K., Nagase, T., Nomura, N., Kikuchi, H., Masuho, Y., Yamashita, R., Nakai, K., Yada, T., Nakamura, Y., Ohara, O., Isogai, T. and Sugano, S.
Complete sequencing and characterization of 21,243 full-length human cDNAs
Nat. Genet. 36 (1), 40-45 (2004)
14702039

TITLE
JOURNAL
PUBMED
REFERENCE
AUTHORS
Isogai, T., Ota, T., Hayashi, K., Sugiyama, T., Otsuki, T., Suzuki, Y., Nishikawa, T., Nagai, K., Sugano, S., Takahashi-Fujii, A., Hara, H., Tanase, T., Nomura, Y., Togiya, S., Komai, F., Hara, R., Takeuchi, K., Arita, M., Nabekura, T., Ishii, S., Kawai, Y., Saito, K., Yamamoto, J., Wakamatsu, A., Nakamura, Y., Nagahara, K., Masuho, Y. and Oshima, A.
NEO human cDNA sequencing project
Unpublished
3 (bases 1 to 2326)
Isogai, T. and Otsuki, T.
Direct Submission
Submitted (23-AUG-2000) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail: genomics@hri.co.jp, Tel: 81-438-52-1975, Fax: 81-438-52-3986)
NEO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology; cDNA library

construction, 5'- and 3'-end one pass sequencing and clone selection:
 Helix Research Institute (supported by Japan Key Technology Center
 etc.) and Department of Virology, Institute of Medical Science,
 University of Tokyo.

FEATURES

source

Location/Qualifiers

1..2326

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="THYRO1001738"

/tissue_type="thyroid gland"

/clone_lib="THYRO1"

/note="cloning vector: pME18SFL3"

94..>2326

/notes="unnamed protein product"

/codon_start=1

/protein_id="BAB14741.1"

/db_xref="GI:10436084"

/translation="MNRLRNKTYVERAVKQKIPTIOGYPVIRCLRRRGWVEKKM
 VHSGLTLPQKDLSSAMGSDTTEDEDEDEFQPSQLDFDLDLKKFDDLDGTH
 ALMGVGLNLRLNLPWFEDVDANSFFRCYCLGABDDKAFIEDFWLTAARNVLKLVVK
 SEWKSYPFOIAVEEASGDKPKQKKNPVLSPFEDFALCACEEYLSLAHMDIDKD
 LEAPLYLTPGSLFLQRYQVQVHEGAEHLHLDTVQRCEDILQQLQAVVQIDMEGD
 RNLIWPKGAKSRGRCIMCMHLEMLKLVNGNPVVMKDGKVVQKYIERPLLIFTGK
 FDLRQWFLVTDNPLTVNWPYRDSYIRFSTQPSLKNLDSVHLKNNISIOKHLNCSHR
 HPLLPPDNWSSORFOALHQEMGAPNAWSTIIVPGMKDAVIHALQTSQTVQCRKASF
 ELYGADFVEGEQFQVLIIEINASPTWAPSTAVTARLCAGVQADTLRWIDRDNCD
 TGAFELIYQPAVEPVQYVIGIRLLVEGFTIKPMANCMHRMGVRPAVPLLTQRGSEA
 RHFPSPSLHTKAOLPSPHVLRHGOVLRHSLKVLGTALSTTGKALRTPLTAKVFLSL
 PNULDPKAPSLILKPKAPALLCLRGPLEVPCCLCPLKEQFAPVGRSPKANSRP
 DCDKPRAEACPMKRLSLPLPLPLVGTGFRRRRLGDMKLGKPLLRFTALVLDPTNKK
 KQVK"

CDS

937 GTGAAGCCAGGAGCCCAAGTCCCGCGGAGCGAGCATCATGTGATGGACCACTGGAGGAG 996
 162 lIleLeuGluLeuAlaAlaAspHisProLeuSerArgAspAsnLysTrpValValGln 181
 997 ATGCTGAAGCTGGTGAACCGCAACCCCGGTGTGATGAAGGACGGCAAGTGGTGGTGCAG 1056
 182 LysTyrIleGluThrProLeuLeuIleCysAspThrLysPheAspIleArgGlnTrpPhe 201
 1057 AAGTATATTGAGCGGCCCTCTCTCATCTTTGCGACCAAGTTTGACTCAGACAGTGGTTC 1116
 202 LeuValThrAspTrpAsnProLeuThrIleTrpPheTyrLysGluSerTyrLeuArgPhe 221
 1117 CTGGTAACCTGACTGGAACCCACTTACCGGTGTGTTCTACCGCGACAGCTATATCCGCTTT 1176
 222 SerThrGlnArgPheSerLeuAspLysLeuAspSerAlaIleHisLeuCysAsnAanAla 241
 1177 TCCAGCGACGCCCTTCTCCCTGAAGAACCTGGCAACTAGTCACCTGTGCAACAACCTCC 1236
 242 ValGlnLysTyrLeuLysAsnAspValGlyArgSerProLeuLeuProAlaHisAsnMet 261
 1237 ATCCAGAACACCTGGAGAACTCATGCGCATCGCATCCACTGCTTCGGCCAGACACAATG 1296
 262 TrpThrSerThrArgPheGlnGluTyrLeuGlnArgGlnGlyArgGlyAlaValTrpGly 281
 1297 TGGTCTAGCCAGAGGTTCCAGGCCACCTGCAGGAGATGGGTGCCCAATGCTTGGTCC 1356
 282 SerValIleTyrProSerMetLysLysAlaIleAlaHisAlaMetLysValAlaGlnAsp 301
 1357 ACCATCATCGCTGCGCATGAAGGATGCTGTATCCACGCACTTCAGACCTCCCAAGGAC 1416
 302 HisValGluProArgLysAsnSerPheGluLeuTyrGlyAlaAspPheValLeuGlyArg 321
 1417 ACCGTGACGTGTCGAAGGCCAGCTTTGAGCTCTATGGCGCTGACTTCGTGTGCGGGAG 1476
 322 AspPheArgProTrpLeuIleGluLysAsnSerSerProThrMetHisProSerThrPro 341
 1477 GACTTCAGCCCTTGGCTGATTGAGATCAACGCCAGCCCGCCAGCATGGACCTCCACAGCA 1536
 342 ValThrAlaGlnLeuCysAlaGlnValGlnGluAspThrIleLysValAlaVal----- 359
 1537 GTCACCTGCCCGCTCTGTGCTGGCGTGAAGCTGACACACCTGCGGTGGTCATACCGCG 1596
 360 -----AspArgSerCysAspIleGlyAsnPheGluLeuTrpArgGlnProValVal 377
 1597 ATGCTGCGCGCAACTGTGACAGAGCGCTTTGAGCTTATCTATATAGCAGCTCTGCTGTG 1656
 378 GluProProProPheSerGlySerAspLeuCysValAlaGlyValSerValArgAla 397
 1657 GAGTGCTCAATATGTGGGCATCCGCTCTGTGTAGAGGGCTTCACCATCAAGAAG--- 1713
 398 ArgArgGlnValLeuProValCysAsnLeuLysAlaSerAlaSerLeuLeuAspAlaGln 417
 1713 ----- 1713
 418 ProLeuLysAlaArgGlyProSerAlaMetProAspProAlaGlnGlyProProSerPro 437
 1714 -----CCATGGCGATG----- 1725
 438 AlaLeuGlnArgAspLeuGlyLeuLysGluLysGlyLeuProLeuAlaLeuLeuAla 457
 1726 ---TGTCATCGCGGATGGGGTCCGC-----CCAGCAGCTCCTCTGCTGT 1767
 458 ProLeuArgGlyAlaAlaGluSerGlyGlyAlaAlaGlnProThrArgThrLysAlaAla 477
 1768 ACCCAGCGAGGCTCTGGGGAAGCCGCTCACCACCTTCCCGACCTCCACCAAGGCC--- 1824
 478 GlyLysValGluLeuProAlaCysProCysArgHisValAspSerGlnAlaProAsnThr 497
 1825 -----CAGCTGCTTCTCC--CCATGTACTCCGACACCGAGGCCGAGTCTCTCAGACG 1874
 498 GlyValProValAlaGlnProAlaLysSerTrp-----AspProAsnGln--- 512

ORIGIN

Alignment Scores:

Pred. No.: 2,88e-46 Length: 2326
 Score: 1054.00 Matches: 233
 Percent Similarity: 55.4% Conservative: 79
 Best Local Similarity: 41.4% Mismatches: 166
 Query Match: 36.9% Indels: 86
 DB: 8 Gaps: 12

US-10-635-977-2 (1-541) x AK023960 (1-2326)

Qy 2 AlasSerIleLeuLysTrpValSerHisGlnSerCysSerArgSerArgSer 21
 Db 544 GCCCGCAACGTTCTCAAGCTGGTGGTGAAGTCTGAG-----TGAAGTCA 588
 Qy 22 LysProArgAspGlnArgGluGluAlaGlySerSerAspLeuSerSerArgGlnAspAla 41
 Db 589 TACCCCTATTACGCGAGTAGAAGAGAGCGCTCAGGAGACAGAGCCCAAGAACAGGAG 648
 Qy 42 GluAsnAlaGluAlaLysLeuArgGlyLeuProGlyGlnLeuValAspIleAlaCysLys 61
 Db 649 AAAAACCCAGTGTG-----GTGTCCCGCAGAGTTTGGGATGAAGCTCTGTGT 696
 Qy 62 ValCysGlnAlaTyrLeuGlyGlnLeuGluHisGluAspIleAspThrSerAlaAspAla 81
 Db 697 GCGTGCAGGAGTACCTTAGCAACTTGGCCACATGGACATCGACAAGGACCTGGAGGCC 756
 Qy 82 ValGluAspLeuThrGluAlaGluTrpGluAspLeuThrGlnGlnTyrTyrSerLeuVal 101
 Db 757 CCGCTGTACTCTACCCCGAGGGCTGGTCCCTCTCTCCAGCGCTACTACCAAGTGGTC 816
 Qy 102 HisGlyAspAlaPheIleSerAsnSerArgAsnTyrPheSerGlnCysGlnAlaLeuLeu 121
 Db 817 CACGAAGGGCAGAACTCAGGCACCTCGACACTCAGGTCCAGCGCTGTGAGACATCTGT 876
 Qy 122 AsnArgIleThrSerValAsnProGlnThrAspIleAspGlyLeuArgAsnIleTrpIle 141
 Db 877 CAGCAGCTGCAGCGCGTGGTACCCAGATAGACATAGCAAGGGGATCGCAACATCTGGATC 936
 Qy 142 IleLysProAlaAlaLysSerArgGlyArgAspIleValCysMetAspArgValGluGlu 161

```

Db 1875 -----ACAGCAGCAGCAGCTGGTGGCAGCTAAGGCCCTGTCTGCACACAGGCAAGGC 1925
QY 513 -----LeuAsnAlaHisProLeuGluProValLeuArg 523
Db 1926 CTTGAGGAGCTTACCCAGCGTCAAGGCTCTTCAATTCCTCCACCGAACCTTGATTTCAA 1985
QY 524 GY-----LeuysThrAlaGluGlyAlaLeuArgPro-----Pro 535
Db 1986 GGTGGCAGCAGATCTGAAGCAAGAAAGGCTCTGTCTCTCTGTGCTCCGAGGCC 2045
QY 536 ProGlyGly 538
Db 2046 CCAGCTGGA 2054

RESULT 9
HSM800637
LOCUS
DEFINITION Homo sapiens mRNA; cDNA DKFp434B103 (from clone DKFp434B103);
partial cds.
ACCESSION AL096725
VERSION AL096725.1 GI:5419858
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1 (bases 1 to 2553)
Wambutt,R., Heubner,D., Mewes,H.W., Gassenhuber,J. and Wiemann,S.
Direct Submission
Submitted (16-JUN-1999) MIPS, Am Klopferspitze 18a, D-82152
Martinsried, GERMANY
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing
consortium of the German Genome Project.
This clone (DKFp434B103) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
information about the clone and the sequencing project is available
at http://www.mips.biochem.mpg.de/proj/cDNA/.

FEATURES
source
1..2553
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="RZPD:DKFp434B103"
/db_xref="taxon:9606"
/clone="DKFp434B103"
/tissue_type="testis"
/clone_lib="434 (synonym: htes3). Vector pSport1; host
DH10B; sites NotI + SalI"
/dev_stage="adult"
1..2553
/gene="DKFp434B103"
921..1950
/gene="DKFp434B103"
/feature="similarity to tubulin-tyrosine ligase"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAB46375.1"
/db_xref="GI:5419859"
/db_xref="GOA:Q9YAR7"
/db_xref="UniProt/Swiss-Prot:Q9YAR7"
/translation="MDI KDLEAPLYLPFGMSLFLQRYOVVHGEALRLDQVOR
CEDILLOQLQAVVPOIDMEGRDNIWVFKAKSRGRCINCMHLEMLKLVNGPVVWK
DGKVVQKIERPLIFGKFDLRFVLTWNELTVWFYRDSYRSTQFNSKLLD.
NSVHLCKNSIQKLENSCHRHPLLPDNNSSQFQALQEMGAPNASTIIVFGMKD.
AVIHALQTSQTVQCKRASFELYGADFVFGDFQFLIEINASPTMAPSTAVTARLCA
GVQADTLRWIDRLMDRNCDTGAFELIYKQPVTTSPASTPRPSPCLLPWYSSTRSSD
DSTASWALRPCRQAPR"
2511..2516
polyA_signal
/gene="DKFp434B103"
2528
polyA_site

```

ORIGIN

```

Alignment Scores:
Pred. No.: 1,91e-44 Length: 2553
Score: 1020.00 Matches: 220
Percent Similarity: 55.5% Conservatives: 69
Best Local Similarity: 42.2% Mismatches: 151
Query Match: 35.7% Indels: 81
DB: 8 Gaps: 12

US-10-635-977-2 (1-541) x HSM800637 (1-2553)
QY 54 GlnLeuValAspIleAlaCysLysValCysGlnAlaTyrLeuGlyGlnLeuGluHisGlu 73
Db 835 GAGTTTGTGGATGAAGCTCTGTGTGCGTGGAGGAGTACCTTAGCAACTTGGCCACATG 894
QY 74 AspIleAspThrSerAlaAspAlaValGluAspLeuThrGluAlaGluTrpGluAspLeu 93
Db 895 GACATCGACAAGGACCTGGAGGCCCTGTGTACCTCACCCCGAGGGCTGGTCTCTTC 954
QY 94 ThrGlnGlnTyrTyrSerLeuValHisGlyAspAlaPheIleSerAsnSerArgAsnTyr 113
Db 955 CTCGAGCGCTACTACCAAGTGTCCACGAAGGGGAGAACTCAGGCACCTCGACACTCAG 1014
QY 114 PheSerGlnCysGlnAlaLeuLeuAsnArgIleThrSerValAsnProGlnThrAspIle 133
Db 1015 GTCCAGCGCTGTGAGGACATCCTGCAGGAGCTGCAGGCGCTGGTACCCAGATAGACATG 1074
QY 134 AspGlyLeuArgAsnIleTrpIleLeuLysProAlaAlaLysSerArgGlyArgAspIle 153
Db 1075 GAAGGGGATCGCAACATCTGGATCTGTAAGCCAGGAGCCAGTCCCGCGAGCGAGCATC 1134
QY 154 ValCysMetAspArgValGluGluLeuLeuAlaAlaAspHisProLeuSer 173
Db 1135 ATGTCATGGACCCACCTCGAGGAGATGCTGAAGCTGGTGAAGCGCAACCCCTGGTGATG 1194
QY 174 ArgAspAsnLysTrpValGlnLysTyrIleGluThrProLeuLeuLeuCysAspThr 193
Db 1195 AAGGAGCGCAAGTGGTGGTGCAGAGAATATATTGAGCGGCCCTCTCTCATCTTTGGCACC 1254
QY 194 LysPheAspIleArgGlnTrpPheLeuValThrAspTrpAsnProLeuThrIleTrpPhe 213
Db 1255 AAGTTTGAACCTCAGACAGTGTCTCTGTGTAACCTGGAACCCACTTACCTGGTGTTC 1314
QY 214 TyrLysGluSerTyrLeuArgPheSerThrGlnArgPheSerLeuAspLysLeuAspSer 233
Db 1315 TACCGGACAGCTATATCCGCTTTTCCAGCGAGCCCTTCTCTCTGAAGAACCTGGACAAAC 1374
QY 234 AlaIleHisLeuCysAsnAsnAlaValGlnLysTyrLeuLysAsnAspValGlyArgSer 253
Db 1375 TCAGTGCACCTGTGCAACAACCTCCATCCAGAGGACCTGGAGAACTCATGCGCATCGGCAT 1434
QY 254 ProLeuLeuProAlaHisAsnMetTrpThrSerThrArgPheGlnGluTyrLeuGlnArg 273
Db 1435 CCAGTGTCTCCGCCAGACACACATGTGTCTAGCCAGAGAGTTCACAGGCCACCTGCAGGAG 1494
QY 274 GlnGlyArgGlyAlaValTrpGlySerValIleTyrProSerMetLysLysAlaIleAla 293
Db 1495 ATGGTGGCCCCAATGTGTGTCACCATCATGTGCTGCGCTGGCATGAAGATCTGTGATC 1554
QY 294 HisAlaMetLysValAlaGlnAspHisValGlnProArgLysAsnSerPheGluLeuTyr 313
Db 1555 CACGCACTTCAGACCTCCAGGACACCGTGCAGTGTCCGAGAGGCCAGCTTTGAGCTCTAT 1614
QY 314 GlyAlaAspPheValLeuGlyArgAspPheArgProTrpLeuIleGluLeuAsnSerSer 333
Db 1615 GCGCTGACTTCGTGTTTCGGGGAGGACTTCCAGCGCTCGCTGATGATGATCAACGCCAGC 1674
QY 334 ProThrMetHisProSerThrProValThrAlaGlnLeuCysAlaGlnValGlnGluAsp 353
Db 1675 CCCAGATGGCACCCCTCCACAGCAGTCACTCCCGGCTCTGTGCTGGCGTGCAGACTGAC 1734

```



```

QY 388 sValAlaGlyValSerValArgAlaArgGlnValLeuProValCysAsnLeuIy 408
Db 2312 -----
QY 408 sAlaSerAlaSerLeuLeuAspAlaGlnProLeuLysAlaArgGlyProSerAlaMetPr 428
Db 2313 ----CCAGTGCCTTCCCTCCCTGACTCGACACAGGCCAGGCTCCACAGCACAGC 2368
QY 428 oAspProAlaGlnGlyProProSerProAlaLeuGlnArgAspLeuGlyLeuLysGluGl 448
Db 2369 ACAGCAAGCTGGTGGGCACTAAGGCCCTGTGCACACAGGCAA---GGCTTGAGGACTC 2425
QY 448 uIys-----GlyLeuProLeuAlaLeuAlaPro---LeuArgGlyAla----- 462
Db 2426 TACCCACGGCTAAGGCTTTCATTTCCCTCCACCGAACCTGATTTCAGAGTGGCACCCA 2485
QY 463 -----AlaGluSerGlyGlyAlaAlaGlnProThrArgThr----- 474
Db 2486 GCATCTGAAGCAAGAAAGGTGGCTCGACTGTGACTCACACCGAGTGCACAGTGTCT 2545
QY 475 -----LysAlaAlaGlyLysValGluLeuProAlaCysProCysAr 488
Db 2546 GAGCAGCGGCTCAGGCTGGAGGCACAGGAGGCGAGCTCCACAGGCTGG----- 2597
QY 488 gHisValAepSerGlnAlaProAsnThrGlyValProValAlaGlnProAlaLysSerTr 508
Db 2598 -----CTGGCACCCCAAGGAGGAGCTG 2620
QY 508 pAspProAenGlnLeuAsn-----AlaHisProLeuGluProValLe 522
Db 2621 GTCTCCCTCAGAGGCCCTTCCTCCACAGACTTCGTGATCATCTCTCTCCCTCTCT 2680
QY 522 uArgGlyLeuLysThrAlaGluGlyAlaLeuArgPro-----ProProGlyGl 538
Db 2681 T-----TCACACCGAGGCTCTGCTCTCTGCTCCGAGGCCCCAGCTGG 2728
QY 538 Y 538
Db 2729 A 2729

RESULT 11
LOCUS BC099735
DEFINITION Homo sapiens tubulin tyrosine ligase-like family, member 3, mRNA
ACCESSION BC099735
VERSION BC099735.1 GI:71043427
KEYWORDS MGC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1769)
AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
Datchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
Carninci,P., Prange,C., Raha,S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Huiyk,S.W.,
Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahey,J., Helton,E., Kettelman,M., Madan,A., Rodrigues,S.,
Sanchez,A., Whitting,M., Madan,A., Young,A.C., Shevchenko,Y.,
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smalls,D.E.,
Schnurich,A., Schein,J.E., Jones,S.J. and Marra,M.A.

```

CONSRMT

Mammalian Gene Collection Program Team

Generation and initial analysis of more than 15,000 full-length

human and mouse cDNA sequences

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

12477932

2 (bases 1 to 1769)

NIH MGC Project

Direct Submission

Submitted (15-JUL-2005)

National Institutes of Health, Mammalian

Gene Collection (MGC), Bethesda, MD 20892-2590, USA

NIH-MGC Project URL: http://mgc.nci.nih.gov

Contact: MGC help desk

Email: cgapbs@mail.nih.gov

Tissue Procurement: Baylor Human Genome Sequencing Center

cDNA Library Preparation: Baylor Human Genome Sequencing Center

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: National Institutes of Health Intramural

Sequencing Center (NISC),

Gaithersburg, Maryland;

Web site: http://www.nisc.nih.gov/

Contact: nisc.mgc@hgrl.nih.gov

Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,

Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,

Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,

Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaepi,R.,

Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C.,

McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W.,

Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,

Young,A., Zhang,L.-H. and Green,E.D.

Clone distribution: MGC clone distribution information can be found

through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov

Series: IRAM Plate: 4 Row: p Column: 10.

Location/Qualifiers

1. .1769

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clones="MGC:120532 IMAGE:40025662"

/tissue_type="PCR rescued clones"

/clone_lib="NIH MGC_283"

/note="vector: PCR-Blunt II-TOPO with reversed insert;"

/note="identification sequence tag: AGTCAG"

1. .1769

/gene="TTLL3"

/note="synonyms: DKFP434B103, HOTTL"

/db_xref="GeneID:26140"

360. .1418

/gene="TTLL3"

/codon_start=1

/product="TTLL3 protein"

/protein_id="AAH99735.1"

/db_xref="GI:71043428"

/db_xref="GeneID:26140"

/translation="MDIDKDLAPLTPGMSLFLQRYQVWHEGASRLHDTQVQR

CEDTLQLOAVVPOIDMEGRDNIWVKGAKSRGIMCMHLEMLKLVGNPVMK

DKGWVQYIKERPLLIIGTRFDLQWFLVTDWNLPTWYFDRSYRSTQPSFLNKLD

NSVHLKNNISQIKLENSCHRRPLLPDNNSSQRFQALQEMGAPNASTIIVPGMKD

AVTHALQSDTQVCRKASFLYGDVFGEDFQFLIEINASPTMAPSTAVTARLCA

GVQADTLRVTDLRNCRDGTGAFELIYKQPVTTSPASTPRPSCILLPMYSIDTRRSSD

DSTASWALRPCRQARP"

ORIGIN

Alignment Scores:

Pred. No.:

Score:

Percent Similarity:

Best Local Similarity:

Query Match:

DB:

1.19e-43

1001.50

56.9%

43.9%

35.1%

8

Length:

Matches:

Conservative:

Mismatch:

Indels:

Gaps:

1769

218

65

142

74

10

US-10-635-977-2 (1-541) x BC099735 (1-1769)

QY	54	GlnLeuValAspIleAlaCysLysValCysGlnAlaTyrLeuGlyGlnLeuGluHisGlu	73	393	SerValArgAlaArgGlnValLeuProValCysAsnLeuLysAlaSerAlaSer	412
DB	303	GAGTTTGTGATGAAGCTCTGTGTGGGTGGAGAGTACCTTAGCAACTTGGCCCATG	362	1383	GCATAAGCCCTGTGCACACAGCAAGGCCTTG-AGGACTCTACCCAGCGTAAAGTC	1441
QY	74	AspIleAspThrSerAlaAspAlaValGluAspLeuThrGluAlaGluTrpGluAspLeu	93	413	LeuLeuAspAlaGlnPro-----LeuLysAlaArg	422
DB	363	GACATCGACAAAGACCTGGAGGCGCCGCTGTACTCACCCCGAGGGCTGGTCCCTCTTC	422	1442	TTTATTTCCTCCCAACCACTTGTATTCAAGTGGCACCAGCATCTCTGAAGCCAAGA	1501
QY	94	ThrGlnGlnTyrTrpSerLeuValHisGlyAspAlaPheIleSerAsnSerArgAsnTyr	113	423	GlyProSerAlaMetProAspProAlaGlnGlyProProSerProAlaLeuGlnArgAsp	442
DB	423	CTCCAGCGCTACTACCAAGTGGTCCAGGAAGGGCAGAACTCAGGCACCTCGACACTCAG	482	1502	AA-GGCTCTGTCTCTGTCTCGAGGCCCGCCAGCTGGAGTGCCTTG-----	1551
QY	114	PheSerGlnCysGlnAlaLeuLeuAsnArgIleThrSerValAsnProGlnThrAspIle	133	443	LeuGlyLeuLysGluLysGlyLeuProLeuAlaLeuLeuAlaProLeuArgGlyAla	462
DB	483	GTCAGCGCTGTGAGGACATCTGCAGCAGCTCGACGGCGGTGTACCCAGATAGACATG	542	1552	-----TTGCCTCTG-----CCC	1563
QY	134	AspGlyLeuArgAsnIleTrrIleIleLysProAlaAlaLysSerArgGlyArgAspIle	153	463	AlaGluSerGlyGlyAlaAlaGlnProThrArg-----ThrLysAlaAlaGlyLysValGlu	481
DB	543	GAAGGGGATCGCAACATCTGGATCGTGAGCCAGGAGCCAAAGTCCCGTGACAGGCAATC	602	1564	TCTGAAGTCGGAACAATTCTAGCACCTGTGCGAAGGTCAAGGCCCAAGCAAAATTCAG	1623
QY	154	ValCysMetAspArgValGluGluIleLeuGluLeuAlaAlaAspHisProLeuSer	173	482	-----LeuProAlaCysProCysArgHisValAsp	491
DB	603	ATGTGATGACACCTGGAGAGATGCTGAAGCTGTGTGAACGCAACCCGCTGGTGATG	662	1624	GCCAGACTGTGACAAACCCAGGCGTGAAGCCTGCC-----CATGAAGAG	1668
QY	174	ArgAspAsnLysTrpValValGlnLysTyrIleGluThrProLeuLeuIleCysAspThr	193	492	SerGlnAlaProAsnThrGlyValProValAlaGlnProAlaLysSerTrp	508
DB	663	AAGGACGGCAAGTGGGTGGTGCAGAAATATTTGAGCGCGCCCTCTCATCTTTTGGCACC	722	1669	GCTGAGCCCTGAAACC-----CCTGCCCTTGTGG	1701
QY	194	LysPheAspIleArgGlnTrpPheLeuValThrAspTrpAsnProLeuThrIleTrpPhe	213	RESULT 12		
DB	723	AAGTTTGACCTCAGACAGTGTCTGTGTAACGTGGAACCCACTTACCGTGTGGTTC	782	BC098361	1743 bp	mRNA
QY	214	TyrLysGluSerTyrLeuArgPheSerThrGlnArgPheSerLeuAspLysLeuAspSer	233	BC098361	linear	PRI 03-AUG-2005
DB	783	TACCGCAGACTATATCCGCTTTCACGACGCCCTTCTCCCTGGAAGACCTGGACAAC	842	BC098361.1	GI:68532603	
QY	234	AlaIleHisLeuCysAsnAlaValGlnLysTyrLeuLysAsnAspValGlyArgSer	253	ACCESSION		
DB	843	TCAGTCACCTGTGCACAACTCCATCCAGAACCATCGAGAACTCATCCCATCGGCAT	902	VERSION		
QY	254	ProLeuLeuProAlaHisAsnMetTrpThrSerThrArgPheGlnGluTyrLeuGlnArg	273	KEYWORDS		
DB	903	CCACTGCTCCGCAGACACATGTGTCTAGCCAGAGGTTCAGGCCCACTTCAGGAG	962	SOURCE		
QY	274	GlnGlyArgGlyAlaValTrpGlySerValIleTyrProSerMetLysLysAlaIleAla	293	ORGANISM		
DB	963	ATGGGTGCCCAAAATGCTTGTCCACCATCATCGTGCCTGCATGAAGGATGCTGTGATC	1022	REFERENCE		
QY	294	HisAlaMetLysValAlaGlnAspHisValGluProArgLysAsnSerPheGluLeuTyr	313	AUTHORS		
DB	1023	CAGCACTTCAGACCTCCAGGACACCGTGCATTCGGAAGGCCAGCTTTGAGCTCAT	1082			
QY	314	GlyAlaAspPheValLeuGlyArgAspPheArgProTrpLeuIleGluLeuAsnSerSer	333			
DB	1083	GGCGCTCAGTCTGTGTGGGGAGGACTTCAGCCCTCGCTGATTGATGATCAAGCCGAGC	1142			
QY	334	ProThrMetHisProSerThrProValThrAlaGlnLeuCysAlaGlnValGlnGluAsp	353	CONSRM		
DB	1143	CCCAGATGGCACCTCCACAGCAGTCACTGCCCGGCTGTGTCGCGTGAAGCTGAC	1202	TITLE		
QY	354	ThrIleLysValAlaVal-----AspArgSerCysAspIleGlyAsnPheGlu	369	JOURNAL		
DB	1203	ACCTGCGCGTGGTCAATTGACCGGAGGCTGACCGCAACTGTGACAGGAGCCCTTGAG	1262	PUBMED		
QY	370	LeuLeuTrpArgGlnProValGluProPro-----	380	AUTHORS		
DB	1263	CTCATCTATAAGACGCCGCTCACCATCTTCCCGAGCCTCCACCAAGCCGAGCTGCTTC	1322	CONSRM		
QY	381	---ProPhe-----SerGlySerAspLeuCysValAlaGlyVal	392	JOURNAL		
DB	1323	CTCCCCATGTACTCGACACCAAGGGCCAGGCTCTCTCAGACGACACGACAGCAAGCTGGT	1382	REMARK		
				COMMENT		

BC098361 1743 bp mRNA linear PRI 03-AUG-2005
 Homo sapiens tubulin tyrosine ligase-like family, member 3, mRNA
 (CDNA clone MGC:120529 IMAGE:40025659), complete cds.

BC098361
 BC098361.1 GI:68532603
 MGC.
 Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Homnidae; Homo.
 1 (bases 1 to 1743)
 Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
 Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
 Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
 Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
 Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
 Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
 Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
 Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
 Abramson, R.D., Mullany, S.J., Bosak, S.A., McEwan, P.J.,
 McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
 Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
 Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
 Fahey, J., Helton, E., Kettelman, M., Madan, A., Young, A.C., Shevchenko, Y.,
 Bouffard, G.C., Blakesley, R.W., Touchman, J.W., Green, E.D.,
 Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
 Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalios, D.E.,
 Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.

Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 12477932
 2 (bases 1 to 1743)

NIH MGC Project
 Direct Submission
 Submitted (24-JUN-2005) National Institutes of Health, Mammalian
 Gene Collection (MGC), Bethesda, MD 20892-2590, USA
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk

Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Baylor Human Genome Sequencing Center
cDNA Library Preparation: Baylor Human Genome Sequencing Center
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILLUMI)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: <http://www.nisc.nih.gov/>
Contact: nisc_mgc@nhgri.nih.gov
Ahtker,N., Aylele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
Blakesley,R.W., Bouffard,G.G., Green,K., Brinkley,C., Brooks,S.,
Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R.,
Maduro,Q.L., Mastello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C.,
McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W.,
Tauregon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
Young,A., Zhang,L.H. and Green,E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAM Plate: 4 Row: 0 Column: 9.

FEATURES
SOURCE

```

source
location/Qualifiers
1. .1743
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:120529 IMAGE:40025659"
/tissue_type="PCR rescued clones"
/clone_lib="NIH_MGC_283"
/note="vector: PCR-Biunt II-TOPO with reversed insert;
clone identification sequence tag: GAAGGCCCTGCTTCG"
1. .1743
gene

```

gene

CDS

ORIGIN

Alignment Scores:					
Pred. No.:	1.32e-43	Length:	1743		
Score:	1000.50	Matches:	218		
Percent Similarity:	56.7%	Conservative:	64		
Best Local Similarity:	43.9%	Mismatches:	143		
Query Match:	35.1%	Indels:	74		
DB:	8	Gaps:	10		

US-10-635-977-2 (1-541) X BC098361 (1-1743)

QY	54	GlnLeuValAspIleAlaCysLysValCysGlnAlaTyrIeuGlyLnLeuGluHisGlu	73
Db	303	GAGTTTGATGAAGACTCTGTGCTGCAGGAGTAGCTACCTTAGCAACTTGGCCCATCG	362
QY	74	AspIleAspThrSerAlaAspAlaValGluAspLeuThrGluAlaGluTrpGluAspLeu	93
Db	363	GACATCGACAAGCACTGGAGGCCCGCTGTACCTCACCCCGAGGGCTGGTCCTCTTTC	422
QY	94	ThrGlnGlnTyrTyrSerLeuValHisGlyAspAlaPheIleSerAsnSerArgAenTyr	113
Db	423	CTTCACGGCTATTACCAAGTAGTGTCACGAAGGGGCGAGAATCAGGCACCTCAG	482

[illegible]

QY	114	PheSerGlnCysGlnAlaLeuLeuAsnArgIleThrSerValAsnProGlnThrAspIle	133
DB	483	GTCCAGCGCTGTGGACATCTCTGCAGCAGCTTGGTGAAGCGTGTTGCCCATAGACATG	542
QY	134	AspGlyLeuArgAsnIleTrpIleIleLysProAlaAlaLysSerArgGlyArgAspIle	153
DB	543	GAAGGGGATCGCAACATCTGATCGTGAAGCAGGAGCCAGTCCTCGTGAGCAGGCATC	602
QY	154	ValCysMetAspArgValGluLeuGluLeuAlaAlaAspHisProLeuSer	173
DB	603	ATGTGATGGACCCTGGAGGAGATGTGAAGCTGGTGAACGGCAACCCCCTGGTGATG	662
QY	174	ArgAspAsnLysTrpValValGlnLysTyrlleGluThrProLeuLeulleCysAspThr	193
DB	663	AAGACCGCAAGTGGTGGTGCAGAAGTATATTGAGCGGCCCTCTCATCTTTGGCAC	722
QY	194	LysPheAspIleArgGlnTrpPheLeuValThrAspTrpAsnProLeuThrIleTrpHe	213
DB	723	ANGITTAGCCTCAGACAGTGGTCTCTGTGTAAGTACTGGAACCCACTTACCGTGTGGTTC	782
QY	214	TyrLysGluSerTyrLeuArgPheSerThrGlnArgPheSerLeuAspLysLeuAspSer	233
DB	783	TACCGCAGACATATACCGCTTTCCACGCGAGCCCTCTCCCTGGAAGNACTGAGCAAC	842
QY	234	AlaIleHisLeuCysAsnAsnAlaValGlnLysTyrLeuLysAsnAspValGlyArgSer	253
DB	843	TCAGTGCACCTGTGCAACAACCTCCATCCAGAAGCACCTGGAGAACTCATGCTCCATCGGCAT	902
QY	254	ProLeuLeuProAlaHisAsnMetTrpThrSerThrArgPheGlnGluTyrLeuGlnArg	273
DB	903	CCACTGCTTCGCCCAGACAACATGTGGTCTAGCCAGAGGTTTCCAGGCCACCTCGCAGGAG	962
QY	274	GlnGlyArgGlyAlaValTrpGlySerValIleTyrProSerMetLysLysAlaIleAla	293
DB	963	ATGGGTGCCCAATGTGTGTGCACCATCATCTGCTGCTGCATGAAGATGCTGTGATC	1022
QY	294	HisAlaMetLysValAlaGlnAspHisValGluProArgLysAsnSerPheGluLeuTyr	313
DB	1023	CACGCATCTAGACCTCCACGACACCGTGCATGTGGAAGGCCAGCTTTGAGCTCTAT	1082
QY	314	GlyAlaAspPheValLeuGlyArgAspPheArgProTrpLeuilleGluIleAsnSerSer	333
DB	1083	GGCGCTCACTTCGTGTTTCGGGAGGAGACTTCCAGCCCTGGCTGATTGAGATCAACGCCACG	1142
QY	334	ProThrMetHisProSerThrProValThrAlaGlnLysCysAlaGlnValGlnGluAsp	353
DB	1143	CCCACGATGGCACCTCCACAGCAGTCACTGCCCGGCTCTGTGCTGGCGTCAAGCTGAC	1202
QY	354	ThrIleLysValAlaVal-----AspArgSerCysAspIleGlyAsnPheGlu	369
DB	1203	ACCTGCGCGTGGTCAITTGACCGGAGGCTGGACCGCAACTGTGACACAGGAGCCTTTGAG	1262
QY	370	LeuLeuTrpArgGlnProValGluProPro-----	380
DB	1263	CTCATCTATAAGCAGCCCGTCACCACTTCCCCAGCCTCCACCAAGGCCCGAGCTGCCTT	1322
QY	381	---ProPhe-----SerGlySerAspLeuCysValAlaGlyVal	392
DB	1323	CTCCCCATGTACTCCGACACCGCGCAGGTCTTCAGACGACACGACAGCAAGAGTGTGGT	1382
QY	393	SerValArgAlaAlaArgArgInValLeuProValCysAsnLeuLysAlaSerAlaSer	412
DB	1383	GCAC TAGGCCCTGTCGACCCACGACGCAAGGCCCTTG-AGGACTCTACCCACGGCTAAGGTC	1441
QY	413	LeuLeuAspAlaGlnPro-----LeuLysAlaArg	422
DB	1442	TTCAITTTCCCTCCACCGAACCTTTGATTTC AAGGTGGCACCCAGCATCTCTGAAGCCAAGA	1501
QY	423	GlyProSerAlaMetProAspProAlaGlnGlyProProSerProAlaLeuGlnArgAsp	442
DB	1502	AA-GGCTCCTGCTCTCTGTGCTTCGAGGCCCGCCAGCTCGAAGTGCTCTG-----	1551
QY	443	LeuGlyLeuLysGluGluLysGlyLeuProLeuAlaLeuLeuAlaProLeuArgGlyVala	462


```

1552 -----TTGCCTCTGCCCTTT----- 1566
Db
463 AlaGluSerGlycylAlaAlaGlnProThrArg---ThrLysAlaAlaGlyLysValGlu 481
QY |||||
1567 ---GAAGTCGAACAATTCTAGCACCTGCGAAGGTCAGAGGCCAAGGCCAAATTCAG 1623
Db
482 -----LeuProAlaCysProCysArgHisValAsp 491
QY |||||
1624 GCCAGACTGTGACAAACCCAGGGCTGAGGCCTGCC-----CATGAAGAG 1668
Db
492 SerGlnAlaProAsnThrGlyValProValAlaGlnProAlaLysSerTrp 508
QY |||||
1669 GCTGAGCCCCCTGAACC-----CCTGCCCTTGTGG 1701
Db

RESULT 13
AX834642
LOCUS
DEFINITION
Sequence 1766 from Patent EP1347046.
ACCESSION
AX834642
VERSION
AX834642.1 GI:39920777
KEYWORDS
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE
1
AUTHORS
Isogai,T., Sugiyama,T., Otsuki,T., Wakamatsu,A., Sato,H., Ishii,S.,
Yamamoto,J.I., Isono,Y., Hio,Y., Otsuka,K., Nagai,K., Irie,R.,
Tamechika,I., Seki,N., Yoshikawa,T., Otsuka,M., Nagahara,K. and
Masuho,Y.
TITLE
Full-length cDNA sequences
JOURNAL
Patent: EP 1347046-A 1766 24-SEP-2003;
Research Association for Biotechnology (JP)
FEATURES
source
1..2848
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN

Alignment Scores:
Pred. No.: 1.74e-42 Length: 2848
Score: 983.50 Matches: 184
Percent Similarity: 70.1% Conservative: 57
Best Local Similarity: 53.5% Mismatches: 96
Query Match: 34.5% Indels: 7
DB: 6 Gaps: 2

US-10-635-977-2 (1-541) x AX834642 (1-2848)
QY 54 GlnLeuValAspIleAlaCysLysValCysGlnAlaTyrLeuGlyGlnLeuGluHisGlu 73
Db 1294 GAGTTTGTAAGAGCTCTGTGCTGCAGAGGAGTACCTTAGCACTTGGCCCATG 1353
QY 74 AspIleAspThrSerAlaAspAlaValGluAspLeuThrGluAlaGluTrpGluAspLeu 93
Db 1354 GACATCGACAAGGACCTGGAGGCCCGCTGTACCTCACCCCGAGGGCTGCTCCTTC 1413
QY 94 ThrGlnGlnTyrTrpSerLeuValHisGlyAspAlaPheIleSerAsnSerArgAsnTyr 113
Db 1414 CTCAGCGCTACTACCAAGTGGTCCAGAGGGGCAAGTCTAGGCACTTCGACACTCAG 1473
QY 114 PheSerGlnCysGlnAlaLeuLeuAsnArgIleThrSerValAsnProGlnThrAspIle 133
Db 1474 GTCCAGCGCTGTGAGGACATCTCGACAGCTGCAGGCGCTGTACCCAGATAGACATG 1533
QY 134 AspGlyLeuArgAsnIleTrpIleLeleProAlaAlaLysSerArgGlyArgAspIle 153
Db 1534 GAAGGGATCGCAACATCTGATCGTGAAGCCAGGAGCCCAAGTCCCTGGACGAGGAT 1593
QY 154 ValCysMetAspArgValGluGluLeuAlaAlaAspHisProLeuSer 173

```

```

1594 ATGTGTCATGACCACTGGAGGAGATGCTGAAGCTGGTGAACGGCAACCCCGTGTGATG 1653
Db
174 ArgAspAsnLysTrpValGlnLysTyrIleGluThrProLeuLeuIleCysAspThr 193
QY |||||
1654 AAGGACGGCAAGTGGTGGTGCAGAAATATATATGAGCGGCCCTCTCTCATCTTTGGCACC 1713
Db
194 LysPheAspIleArgGlnTrpPheLeuValThrAspTrpAsnProLeuThrIleTrpPhe 213
QY |||||
1714 AAGTTTGACCTCAGACAGTGGTTCCTGGTAACTGACTGGAACCCACTTACCGTGTGTTTC 1773
Db
214 TyrLysGluSerTyrLeuArgPheSerThrGlnArgPheSerIleuAspLysLeuAspSer 233
QY |||||
1774 TACCGCGACAGCTATATCCGCTTTTCCAGCGAGCCCTTCTCCCTGAAGAACCTGGCAAC 1833
Db
234 AlaIleHisLeuCysAsnAsnAlaValGlnLysTyrLeuLysAsnAspValGlyArgSer 253
QY |||||
1834 TCAGTGCACCTGTGCACAACTCCATCCAGAAACACCTGGAGAACTCATGCCATCGGCAT 1893
Db
254 ProLeuLeuProAlaHisAsnMetTrpThrSerThrArgPheGlnGluTyrLeuGlnArg 273
QY |||||
1894 CCACCTCTCCGCCAGACAAACATGGTCTAGCCAGAGGTTCCAGGCCCACTTCAGGAG 1953
Db
274 GlnGlyArgGlyAlaValTrpGlySerValIleTyrProSerMetLysLysAlaIleAla 293
QY |||||
1954 ATGGGTGCCCAAAATGCTTGGTCCACCATCATCTGCTGGCATGAAGGATGCTGTGATC 2013
Db
294 HisAlaMetLysValAlaGlnAspHisValGluProArgLysAsnSerPheGluLeuTyr 313
QY |||||
2014 CAGCGACTTCAGACCTCCCGAGGACACCTGCGCAATTCGGAAGCCAGCTTTGAGCTCTAT 2073
Db
314 GlyAlaAspPheValLeuGlyArgAspPheArgProTrpLeuIleGluLeuAsnSerSer 333
QY |||||
2074 GCGCTGACTTCGTTCGGGTGGACTTCCAGCCCTGCTGATGATGAGATCAACGCCAGC 2133
Db
334 ProThrMetHisProSerThrProValThrAlaGlnLeuCysAlaGlnValGlnGluAsp 353
QY |||||
2134 CCCAGATGGCACCTCCACAGCAGTCACTGCCCGGCTCTGTCTGGTGGTGAAGTGCAC 2193
Db
354 ThrIleLysValAlaVal-----AspArgSerCysAspIleGlyAsnPheGlu 369
QY |||||
2194 ACCCTGGCGTGTGTCATTGACCGAGGCTGGACCGCAACTGTGCACAGAGAGCTTTGAG 2253
Db
370 LeuLeuTrpArgGln-----ProValValGluProProPheSerGlySerAsp 386
QY |||||
2254 CTCATCTAAGCAGGCGCCAGCTCTTAACATGCAGGTGTCCACAGAGAAATGCACCA 2313
Db
387 LeuCysValAla 390
QY |||||
2314 CTGTGCCAGCA 2325
Db

RESULT 14
AX097236
LOCUS
DEFINITION
Homo sapiens cDNA FLJ39917 fis, clone SPLN2019405, highly similar
to Homo sapiens HOT1L protein mRNA.
ACCESSION
AX097236
VERSION
AK097236.1 GI:21756925
KEYWORDS
oligo capping; fis (full insert sequence).
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE
1
AUTHORS
Ota,T., Suzuki,Y., Nishikawa,T., Otsuki,T., Sugiyama,T., Irie,R.,
Wakamatsu,A., Hayashi,K., Sato,H., Nagai,K., Kimura,K., Makita,H.,
Sekine,M., Oyayashi,M., Nishi,T., Shibahara,T., Tanaka,T.,
Ishii,S., Yamamoto,J., Saito,K., Kawai,Y., Isono,Y., Nakamura,Y.,
Nagahara,K., Murakami,K., Yasuda,T., Iwayanagi,T., Wagatsuma,M.,
Shiratori,A., Sudo,H., Hosoiri,T., Kaku,Y., Kodaira,H., Kondo,H.,
Sugawara,M., Takahashi,M., Kanda,K., Yokoi,T., Furuya,T.,
Kikkawa,E., Omura,Y., Abe,K., Kamiyama,K., Katsuta,N., Sato,K.,

```


Tanikawa, M., Yamazaki, M., Ninomiya, K., Ishibashi, T., Yamashita, H., Murakawa, K., Fujimori, K., Tanai, H., Kimata, M., Watanabe, M., Hirooka, S., Chiba, Y., Ishida, S., Ono, Y., Takiguchi, S., Watanabe, S., Yosida, M., Houta, T., Kusano, J., Kanehori, K., Takahashi-Fujii, A., Hara, H., Tanase, T., Nomura, Y., Togiya, S., Komai, F., Hara, R., Takeuchi, K., Arita, M., Imose, N., Musashino, K., Yuuki, H., Oshima, A., Sasaki, N., Aotsuka, S., Yoshikawa, Y., Matsuura, H., Ichihara, T., Shihohata, N., Sano, S., Moriya, S., Momiyama, H., Satoh, N., Takami, S., Terashima, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mizoguchi, H., Goto, Y., Shimizu, F., Wakebe, H., Hishigaki, H., Watanabe, T., Sugiyama, A., Takemoto, M., Kawakami, B., Yamazaki, M., Watanabe, K., Kumagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M., Tashiro, H., Tanigami, A., Fujiwara, T., Ono, T., Yamada, K., Fujii, Y., Ozaki, K., Hiroo, M., Ohmori, Y., Kawabata, A., Hiki, J., T., Kobatake, N., Inagaki, H., Ikema, Y., Okamoto, S., Okitani, R., Kawakami, T., Noguchi, S., Itoh, T., Shigetani, K., Senba, T., Matsuura, K., Nakajima, Y., Mizuno, T., Morinaga, M., Sasaki, M., Togaishi, T., Oyama, M., Hata, H., Watanabe, M., Komatsu, T., Mizushima-Sugano, J., Sato, T., Shirai, Y., Takahashi, Y., Nakagawa, K., Okumura, K., Nagase, T., Nomura, N., Kikuchi, H., Masuho, Y., Yamashita, R., Nakai, K., Yada, T., Nakamura, Y., Ohara, O., Isogai, T. and Sugano, S. Complete sequencing and characterization of 21,243 full-length human cDNAs

JOURNAL
PUBMED
REFERENCE
AUTHORS

Nat. Genet. 36 (1), 40-45 (2004)
14702039

Ninomiya, K., Wagatsuma, M., Kanda, K., Kondo, H., Yokoi, T., Kodaira, H., Furuwa, T., Takahashi, M., Kikkawa, E., Omura, Y., Abe, K., Kamiyama, K., Katsuta, N., Sato, K., Tanikawa, M., Yamazaki, M., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Ota, T., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Murakawa, K., Kanehori, K., Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahara, K., Masuho, Y., Nagai, K. and Isogai, T. NEDO human cDNA sequencing project

Unpublished

3 (bases 1 to 2848)
Isogai, T. and Yamamoto, J.
Direct Submission

Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan (Fax: 81-438-52-3986) (E-mail: genomics@ri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986) NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.

FEATURES
source
Location/Qualifiers
1..2848
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="SPLN2019405"
/tissue type="spleen"
/clone_lib="SPLN2"
/note="Cloning vector: pME18SFL3"

ORIGIN
Alignment Scores:
Pred. No.: 1.74e-42 Length: 2848
Score: 983.50 Matches: 184
Percent Similarity: 70.1% Conservative: 57
Best Local Similarity: 53.5% Mismatches: 96
Query Match: 34.5% Indels: 7
DB: 8 Gaps: 2

US-10-635-977-2 (1-541) x AK097236 (1-2848)

Qy 54 GlnLeuValAspIleAlaCysLysValCysGlnAlaTyrLeuGlnLeuHisGlu 73

Db 1294 GAGTTTGTGGATGAAGCTCTGTGTGGTGGAGGAGTACCTTAGCAACCTTGCCACATG 1353
Qy 74 AspIleAspThrSerAlaAspAlaValGluAspLeuThrGluAlaGluThrGluAspLeu 93
Db 1354 GACATCGAACAGGACCTGGAGCCCGCTGTACCTCACCCCGGAGGGCTGGTCCCTTTC 1413
Qy 94 ThrGlnGlnTyrTyrSerLeuValHisGlyAspAlaPheIleSerAsnSerArgAsnTyr 113
Db 1414 CTCACGCGCTACTACCAAGTGTCTCCACGAAGGGGAGAACTCAGGCACCTCGACACTCAG 1473
Qy 114 PheSerGlnCysGlnAlaLeuLeuAsnArgIleThrSerValAsnProGlnThrAspIle 133
Db 1474 GTCCACGCGCTGTGAGGACATCTCTGCAGCAGCTGCAGGCGCTGTACCCCGGAGGAGT 1533
Qy 134 AspGlyLeuArgAsnIleTyrPheIleLysProAlaAlaLysSerArgGlyArgAspIle 153
Db 1534 GAAGGGATCGCAACATCTGGATCTGGAAGCAGGAGCAAGTCCCGTGGAGAGGCATT 1593
Qy 154 ValCysMetAspArgValGluGluLeuLeuAlaAlaAspHisProLeuSer 173
Db 1594 ATGTGATGGACCACTCGAGGAGATGTGAAGCTGGTGAACGGCAACCCCGTGTGATG 1653
Qy 174 ArgAspAsnLysTyrValValGlnLysTyrIleGluThrProLeuLeuIleCysAspThr 193
Db 1654 AAGGACGGCAAGTGGTGGTGCAGAAAGTATATTGAGCGCGCCCTCTCTATCTTTGGCACC 1713
Qy 194 LysPheAspIleArgGlnTyrPheLeuValThrAspTyrAsnProLeuThrIleTyrPhe 213
Db 1714 AAGTTTGACCTCAGACAGTGGTCTCTGTGTAAGTACTGCTGGAACCCACTTACCGCTGGTTC 1773
Qy 214 TyrLysGluSerTyrLeuArgPheSerThrGlnArgPheSerLeuAspLysLeuAspSer 233
Db 1774 TACCGGACAGCTATATCCGCTTTTCCACGCGAGCCCTTCTCCCTGAAGAACCTGGACAC 1833
Qy 234 AlaIleHisLeuCysAsnAlaValGlnLysTyrLeuLysAsnAspValGlyArgSer 253
Db 1834 TCAGTGACCTGTGTCAACCACTCCATCCAGAGCACTCGAGAACTCATGTCATCGGCAT 1893
Qy 254 ProLeuLeuProAlaHisAsnMetTyrThrSerThrArgPheGlnGluTyrLeuGlnArg 273
Db 1894 CCATGCTTCCGCCAGCAACATGTGTGTCTAGCCAGAGGTTCCAGGCCCACTCGCAGAG 1953
Qy 274 GlnGlyArgGlyAlaValTyrPheLysValIleTyrProSerMetLysLysAlaIleAla 293
Db 1954 ATGGGTGCCCCAAATGCTTGTGTCCACCATCATCTGCTGCATGAGAGATGCTGTGATC 2013
Qy 294 HisAlaMetLysValAlaGlnAspHisValGluProArgLysAsnSerPheGluLeuTyr 313
Db 2014 CACGCACTTCAGACCTCCAGGACACCGTGCATGTCCGAAGGCCAGCTTTGAGCTCTAT 2073
Qy 314 GlyAlaAspPheValLeuLysArgAspPheArgProTyrPheIleGluIleAsnSerSer 333
Db 2074 GCGCTGACTTCTGTTGGGGTGGACTTCCAGCCCTGGCTGATGTAGATCAACGCCAGC 2133
Qy 334 ProThrMetHisProSerThrProValThrAlaGlnLeuCysAlaGlnValGlnGluAsp 353
Db 2134 CCCACGATGGACCCCTCCACAGCAGTCACTGCCCGGCTCTGTGTGGCTGCAAGCTGAC 2193
Qy 354 ThrIleLysValAlaVal-----AspArgSerCysAspIleGlyAsnPheGlu 369
Db 2194 ACCCTGCGCGTGGTCAATTGACCGGAGGCTGGACCGCAACTGTGACAGGAGCGCTTTGAG 2253
Qy 370 LeuLeuTyrArgGln-----ProValValGluProProPheSerGlySerAsp 386
Db 2254 CTCATCTATAGCAGGCGGCCAGCTCTCTAAATGAGGTGTCCACGAGAGAAATGCACCA 2313
Qy 387 LeuCysValAla 390
Db 2314 CTGTGCCAGCA 2325

RESULT 15
BT021865


```
Db 998 GCACCTCCACGGCTGTCACCGCAGGCTCTGTGCCGGTGTGTGAGGCCGACACCCCTGCGT 1057
Qy 357 ValAlaVal-----AspArgSerCysAspIleGlyAsnPheGluLeuLeuTrp 372
Db 1058 GTGGTCATGACTGGGACTGGAGCCGCAC-TGTGACACGGGGCCTTTGAGCTCATCTAC 1116
Qy 373 ArgGlnProValValGluProProPheSerGlySerAspLeuCysValAlaGlyVal 392
Db 1117 AAGCAGCCTGCGGTGAGGTGCCCGCCAGTATGTGGGTATCCGGCTCTAGTAGAGGCTCC 1176
Qy 393 SerValArgAlaArgAlaArgGlnValLeuProValCysAsnLeuLysAlaSerAlaSer 412
Db 1177 ACCATCAAGAG-----CCC 1191
Qy 413 LeuLeuAspAlaGlnProLeuLysAlaArgGlyProSerAlaMetProAspProAlaGln 432
Db 1192 CTGGC-GATGAGTCACCGCGGATGCCCTCGCCCGCAGC----- 1229
Qy 433 GlyProProSerProAla-LeuGlnArgAspLeuGlyLeuLysGluGluLysGlyLeuPr 452
Db 1230 ---CCTCCTCACCTGCTGCGCCCGAGAGGCTCTGGGAAAGCAAGGACTCAGGAAGCCC 1286
Qy 452 oLeu----- 453
Db 1287 CATCCACAGCCCTTACCACCTTCCCAAGCCTCCACAGCAAGCCCGGCTGCGCTTCTCCCC 1346
Qy 454 -AlaLeuLeuAlaProLeuArgGly-----AlaAlaGluSerGlyGlyAlaAl 469
Db 1347 ATGTGCCCTGGCCCCCAGGGCGCAGGTCTCTCAGACTACAGCACAGCAAGCTGGTGGGCACTA 1406
Qy 469 aGlnProThr-----ArgThrLysAlaAlaGlyLysValG1 481
Db 1407 AGGCCCTGTGACCAACAGAGGCTTGATGACTCTACCAACCGCCCAAGGTTTGTATTT 1466
Qy 481 uLeuProAlaCysProCysArgHisValAspSerGlnAlaProAsnThrGlyValProVa 501
Db 1467 CCCTCCACCTAACCTTGAGCTCAGCTGGCACCACCGCCCTGAACCCAAAAAAGGCTC 1526
Qy 501 lAlaGlnProAlaLysSerTrp 508
Db 1527 TTGCTCCCCAGCGCCTTCATGG 1548
```

Search completed: April 4, 2006, 13:37:14
Job time : 7987 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7

Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2006, 12:59:01 ; Search time 5.39777 Seconds
(without alignments)
1364.812 Million cell updates/sec

Title: US-10-635-977-2_COPY_133_374

Perfect score: 1299

Sequence: 1 IDGLNIIWPKAKSRGRD.....IKVAVDRSCDIGNFELLWRQ 242

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 180808 seqs, 30441898 residues

Total number of hits satisfying chosen parameters: 180808

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA New:

- 1: /SIDSS/prodata/2/pubpaa/US08_NEW_PUB.pep.*
- 2: /SIDSS/prodata/2/pubpaa/US06_NEW_PUB.pep.*
- 3: /SIDSS/prodata/2/pubpaa/US07_NEW_PUB.pep.*
- 4: /SIDSS/prodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 5: /SIDSS/prodata/2/pubpaa/US09_NEW_PUB.pep.*
- 6: /SIDSS/prodata/2/pubpaa/US10_NEW_PUB.pep.*
- 7: /SIDSS/prodata/2/pubpaa/US11_NEW_PUB.pep.*
- 8: /SIDSS/prodata/2/pubpaa/US60_NEW_PUB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	271	20.9	439	7	US-11-072-512-3658
2	270	20.8	592	7	US-11-072-512-3371
3	177	13.6	404	7	US-11-072-512-3621
4	79.5	6.1	433	7	US-11-167-836-26
5	79	6.1	462	7	US-11-096-568A-25994
6	79	6.1	470	7	US-11-096-568A-25993
7	79	6.1	517	7	US-11-096-568A-25992
8	79	6.1	549	7	US-11-085-185-2
9	78	6.0	256	7	US-11-167-831-9
10	78	6.0	256	7	US-11-167-831-14
11	78	6.0	667	6	US-10-793-626-198
12	77	5.9	256	5	US-09-810-501-67
13	77	5.9	256	7	US-11-167-831-8
14	77	5.9	256	7	US-11-167-831-11
15	76	5.9	256	7	US-11-167-831-12
16	76	5.9	256	7	US-11-167-831-13
17	75	5.8	362	7	US-11-012-762-56
18	75	5.8	1047	7	US-11-072-512-2408
19	75	5.8	1137	7	US-11-087-099-5799
20	74.5	5.7	646	7	US-11-087-099-10725
21	74.5	5.7	698	7	US-11-087-099-8952
22	74.5	5.7	932	7	US-11-017-550-65
23	74.5	5.7	932	7	US-11-002-387-65
24	74	5.7	179	7	US-11-072-512-2871
25	74	5.7	255	7	US-11-167-831-10

26	73	5.6	351	7	US-11-096-568A-5676	Sequence 5676, Ap
27	73	5.6	436	7	US-11-096-568A-5675	Sequence 5675, Ap
28	73	5.6	497	6	US-10-793-626-2812	Sequence 2812, Ap
29	73	5.6	729	6	US-10-995-561-878	Sequence 878, App
30	73	5.6	744	6	US-10-995-561-876	Sequence 876, App
31	73	5.6	752	6	US-10-995-561-879	Sequence 879, App
32	73	5.6	753	6	US-10-995-561-877	Sequence 877, App
33	73	5.6	1131	7	US-11-087-099-2209	Sequence 2209, Ap
34	73	5.6	1131	7	US-11-087-099-9522	Sequence 9522, Ap
35	72.5	5.6	302	7	US-11-098-686-10927	Sequence 10927, A
36	72.5	5.6	698	7	US-11-087-099-9341	Sequence 9341, Ap
37	72.5	5.6	888	6	US-10-131-826A-544	Sequence 544, App
38	72.5	5.6	888	6	US-10-973-115B-544	Sequence 544, App
39	71.5	5.5	420	7	US-11-096-568A-32310	Sequence 32310, A
40	71.5	5.5	508	7	US-11-096-568A-32309	Sequence 32309, A
41	71.5	5.5	508	7	US-11-172-740-259	Sequence 259, App
42	71.5	5.5	508	7	US-11-172-740-300	Sequence 300, App
43	71.5	5.5	512	7	US-11-096-568A-32308	Sequence 32308, A
44	71	5.5	471	7	US-11-024-959-399	Sequence 399, App
45	70.5	5.4	121	7	US-11-107-028-16	Sequence 16, Appl

ALIGNMENTS

RESULT 1

US-11-072-512-3658
; Sequence 3658, Application US/11072512
; Publication No. US20060029945A1

; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO

; APPLICANT: SUGIYAMA, TOMOYASU

; APPLICANT: OTSUKI, TETSUJI

; APPLICANT: WAKAMATSU, AI

; APPLICANT: SATO, HIROYUKI

; APPLICANT: ISHII, SHIZUKO

; APPLICANT: YAMAMOTO, JUN-ICHI

; APPLICANT: ISONO, YUUKO

; APPLICANT: HIO, YURI

; APPLICANT: OTSUKA, KAORU

; APPLICANT: NAGAI, KEIICHI

; APPLICANT: IRIE, RYOTARO

; APPLICANT: TAMECHIKA, ICHIRO

; APPLICANT: SEKI, NAOHICO

; APPLICANT: YOSHIKAWA, TSUTOMU

; APPLICANT: OTSUKA, MOTOKUKI

; APPLICANT: NAGAHARI, KENJI

; APPLICANT: MASUHO, YASUHIKO

; TITLE OF INVENTION: Novel full length cDNA

; FILE REFERENCE: 084335-0191

; CURRENT APPLICATION NUMBER: US/11/072,512

; CURRENT FILING DATE: 2005-03-07

; PRIOR APPLICATION NUMBER: US 60/350,978

; PRIOR FILING DATE: 2002-01-25

; PRIOR APPLICATION NUMBER: JP 2001-379298

; NUMBER OF SEQ ID NOS: 4096

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 3658

; LENGTH: 439

; TYPE: PRT

; ORGANISM: Homo sapiens

US-11-072-512-3658

Query Match 20.9%; Score 271; DB 7; Length 439;

Best Local Similarity 29.8%; Pred. No. 6.3e-20;

Matches 75; Conservative 45; Mismatches 86; Indels 46; Gaps 9;

Qy 8 WIKPAKSRGRDVCMDRVEEILEAAADHPLSRDNK-----WVQKVIETPLICD 60

Db 164 WIMKPVARSOKGLFFLRLLKIDVDM-RKOTRRSDQDDIPVNYAQRVIENPYLIGG 222

Qy 61 TKFDIRQWFLVTWNPLTIWFYKESYLRFSTQRFSLDKLDSAIHLCNNAVQKYLKNDVGR 120


```
QY 60 DTKFDIRQWFLVTDWNL-----TIWFKESYL-----FSTQRFSLDKLDSAIHLCCNAV 110
Db 93 DAAFSRPRFVAGKYTAYDCSGLWMAFFEPYLRQARRICATELFSATRLUESFHIIRDEEV 152
QY 111 Q---KYLKNDVGRSPLLPAAHNMWTSRFOYLQQRGAVGWGVIYVPSMKKAIAHAMKVA 167
Db 153 RVMLRQLRQAAGR-----TVRLDYLOMLALGVISRLV---GKK---YVMEEA 195
QY 168 QD-----HVEPRK-----NSFELYGADFVLGRDRFPMW 195
Db 196 ADGEDGSAPAITPAEFREMWDEFFALHGA-FNIG-DYIPWL 234

RESULT 8
US-11-085-185-2
; Sequence 2, Application US/11085185
; Publication No. US20050287634A1
; GENERAL INFORMATION:
; APPLICANT: MACHIDA, Masayuki
; TITLE OF INVENTION: NOVEL GENES OF CELL WALL-DEGRADING ENZYMES DERIVED FROM ASPERGILL
; TITLE OF INVENTION: METHOD FOR THE PRODUCTION OF THE ENZYMES
; FILE REFERENCE: 4600-0110PUS1
; CURRENT APPLICATION NUMBER: US/11/085,185
; CURRENT FILING DATE: 2005-03-22
; PRIOR APPLICATION NUMBER: JP 2004-188849
; PRIOR FILING DATE: 2004-06-25
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 549
; TYPE: PRT
; ORGANISM: Aspergillus oryzae
US-11-085-185-2

Query Match 6.1%; Score 79; DB 7; Length 549;
Best Local Similarity 21.2%; Pred. No. 4.4; Indels 34; Gaps 6;
Matches 38; Conservative 28; Mismatches 79;

QY 27 VEEILELAAADHPLSRDNKMWVKY-IETPLICDTKFDIRQWFLVTDWNLPTIWFYKES 85
Db 323 VEDVLN-----WTTDATKLQKNGISAVTVVHDGFLNLSKWKQMLKDRPDRMFLDTHQ 375
QY 86 YLRSTQRFSLDKLDSAIHLGN---NAVQYKLVNDVGRSPLLPAAHNMWTSRFOEYLQQR 142
Db 376 YTIENTGQIVLNHTDRVKLICNDWYNNMIKEINTTSAGWGPTICGWSQADTDCAQYLLNV 435
QY 143 GRGAVW-----GSVIV-----PSMKKAIAHA-----MKVAQDHVEPRKNSF 178
Db 436 GRGTRWEGTFAIGDSTVCPTADTGPTCSCASANAPPADYSDGYKKFLQTYAEAQMSAF 494

RESULT 9
US-11-167-831-9
; Sequence 9, Application US/11167831
; Publication No. US20060062805A1
; GENERAL INFORMATION:
; APPLICANT: PAUL DR., PREM S
; APPLICANT: ZHANG, YANJIN
; TITLE OF INVENTION: PROTEINS ENCODED BY POLYNUCLEIC ACIDS OF PORCINE
; TITLE OF INVENTION: REPRODUCTIVE AND RESPIRATORY SYNDROME VIRUS (PPRSV)
; FILE REFERENCE: 8199-0005-55XCIP WO
; CURRENT APPLICATION NUMBER: US/11/167,831
; CURRENT FILING DATE: 2005-06-28
; PRIOR APPLICATION NUMBER: US/10/428,826
; PRIOR FILING DATE: 2003-05-05
; PRIOR APPLICATION NUMBER: US/09/601,326
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: PCT/US99/02630
; PRIOR FILING DATE: 1999-04-19
; PRIOR APPLICATION NUMBER: US 09/019,793
; PRIOR FILING DATE: 1998-02-06
; PRIOR APPLICATION NUMBER: US 08/478,316
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/301,435
; PRIOR FILING DATE: 1994-09-01
; PRIOR APPLICATION NUMBER: US 08/131,625
; PRIOR FILING DATE: 1993-10-05
; PRIOR APPLICATION NUMBER: US 07/969,071
; PRIOR FILING DATE: 1992-10-30
; NUMBER OF SEQ ID NOS: 175
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 256
; TYPE: PRT
; ORGANISM: Porcine reproductive and respiratory syndrome virus
```

```
; PRIOR APPLICATION NUMBER: US 08/301,435
; PRIOR FILING DATE: 1994-09-01
; PRIOR APPLICATION NUMBER: US 08/131,625
; PRIOR FILING DATE: 1993-10-05
; PRIOR APPLICATION NUMBER: US 07/969,071
; PRIOR FILING DATE: 1992-10-30
; NUMBER OF SEQ ID NOS: 175
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 256
; TYPE: PRT
; ORGANISM: Porcine reproductive and respiratory syndrome virus
US-11-167-831-9

Query Match 6.0%; Score 78; DB 7; Length 256;
Best Local Similarity 18.8%; Pred. No. 2.1;
Matches 48; Conservative 34; Mismatches 73; Indels 100; Gaps 12;

QY 46 WVVKYIETPLLI-----CDTKFDIRQWFLVTDW-----NPLTIWFKESYL 88
Db 18 WMLSRSSWCPLLISLFWFPCLASPSVGVWGSFASDFAPRYSVRALPFTLSNVRYSYA 77
QY 89 FSTQRFSLDKLDSAIHLCCNAVQKY-LKNDVGRSPLLPAAHNMWTSRFOEYLQQR 141
Db 78 FLSQ-----CQVDIPTGTGKHPG---MLWHKVTSLIDEMVSRMYRIMEK 121
QY 142 QRGAVWGSVIYP-----SMKKAIAHAMKVAQDHVE-----PRKNSFELYGADFV 186
Db 122 AGQ-AAWKQVSEATLSRISLSDVAHFQHLAAIEAEATCKYLASRLFLMHLNLRMTGSNVT 180
QY 187 LG-----RDFRPWLIEINSSTPMHPSTPVTVAQLCAQVQEDTIKV 225
Db 181 IVYNSTLNQVFAIFPTGSRPKLHDFQQLIAVHSS-----IFS 219
QY 226 AVDRSCDIGNFELLW 240
Db 220 SVAASCTL--FVVLW 232

RESULT 10
US-11-167-831-14
; Sequence 14, Application US/11167831
; Publication No. US20060062805A1
; GENERAL INFORMATION:
; APPLICANT: PAUL DR., PREM S
; APPLICANT: ZHANG, YANJIN
; TITLE OF INVENTION: PROTEINS ENCODED BY POLYNUCLEIC ACIDS OF PORCINE
; TITLE OF INVENTION: REPRODUCTIVE AND RESPIRATORY SYNDROME VIRUS (PPRSV)
; FILE REFERENCE: 8199-0005-55XCIP WO
; CURRENT APPLICATION NUMBER: US/11/167,831
; CURRENT FILING DATE: 2005-06-28
; PRIOR APPLICATION NUMBER: US/10/428,826
; PRIOR FILING DATE: 2003-05-05
; PRIOR APPLICATION NUMBER: US/09/601,326
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: PCT/US99/02630
; PRIOR FILING DATE: 1999-04-19
; PRIOR APPLICATION NUMBER: US 09/019,793
; PRIOR FILING DATE: 1998-02-06
; PRIOR APPLICATION NUMBER: US 08/478,316
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/301,435
; PRIOR FILING DATE: 1994-09-01
; PRIOR APPLICATION NUMBER: US 08/131,625
; PRIOR FILING DATE: 1993-10-05
; PRIOR APPLICATION NUMBER: US 07/969,071
; PRIOR FILING DATE: 1992-10-30
; NUMBER OF SEQ ID NOS: 175
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 256
; TYPE: PRT
; ORGANISM: Porcine reproductive and respiratory syndrome virus
```

US-11-167-831-14

Query Match 6.0%; Score 78; DB 7; Length 256;
Best Local Similarity 18.8%; Pred. No. 2.1;
Matches 48; Conservative 34; Mismatches 73; Indels 100; Gaps 12;

QY 46 WVQKYEITPLLI-----CDTKFDIRQWFLVTDW-----NPLTIWFYKESYLR 88
DB 18 WMLSRSSWCLLISLYFWPFCCLASPSQVGMWSPASDFAPRYSVRALPFTLSNYSRYEA 77
QY 89 FSTQRFSLDKLDSAIHLCCNNAVQKY-LKNDVGRSPLLPAHNMWT-----STRFQYLOR 141
DB 78 FLSQ-----CQVDIPTWGTGKPLG---MLMHHKVSTLIDEMVSRMYRIMEK 121
QY 142 QGRGAVGMSVIVP-----SMKKAIAHAMKVAQDHVE-----PRKNSFELYGADFV 186
DB 122 AGQ-AAWKQVSEATLSRISLSDVVAHFQHLAAIEAETCKYLASRLPMLHLRMTGSNVT 180
QY 187 LG-----RDRPWLIEINSSPTMHPSTPVTVAQLCAQVQEDTIKV 225
DB 181 IVYNSTLNQVFAIFPTGSRPKLHDFQWLIAVHSS-----IFS 219
QY 226 AVDRSCDIGNFELLW 240
DB 220 SVAASCTL--FVVLW 232

RESULT 11

US-10-793-626-198
; Sequence 198, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 198
; LENGTH: 667
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence

US-10-793-626-198

Query Match 6.0%; Score 78; DB 6; Length 667;
Best Local Similarity 24.0%; Pred. No. 7.2;
Matches 41; Conservative 34; Mismatches 70; Indels 26; Gaps 9;

QY 73 DWN-----PLTIWFYKESYLRSTQRFSLDKLDSAIHLCCNNAVQKYLRNDVGRSPLLPAH 128
DB 15 EWNLEDDLPIPKQMKTVISKTNRQF---KIVKSLDKNVKEIIATDAGREGELVARL 71
QY 129 MWTSTRFQYLORQGRGAVGMSVIVPSMKKAIAHAMKVAQDHVEPRKNSFE-LYCADFVL 187
DB 72 ILDKVGNKKPIKR-----LWISSV---TKKAIQEGFKQKN-----GNAYQNLVEA--AL 116
QY 188 GRDRPWLIEINSSPTMHPSTPVTVAQL-CAQVQEDTIKVAVDRSCDIGNFE 237
DB 117 ARSEADWIVGINATRAL--TTYDAQLSLGRVQTFTIQIVKSQDEINFK 165

RESULT 12

US-09-810-501-67
; Sequence 67, Application US/09810501
; Publication No. US20060029928A1
; GENERAL INFORMATION:
; APPLICANT: PAUL, PREM S.

MENG, XIANG-JIN
HALBUR, PATRICK G.
MOROZOV, IGOR
LUM, MELISSA A.
TITLE OF INVENTION: A POLYNUCLEIC ACID ISOLATED FROM A
PORCINE REPRODUCTIVE AND RESPIRATORY SYNDROME VIRUS (PRRS)
A PROTEIN ENCODED BY THE POLYNUCLEIC ACID, A VACCINE
PREPARED FROM OR CONTAINING THE POLYNUCLEIC ACID OR
NUMBER OF SEQUENCES: 77
CORRESPONDENCE ADDRESS:
ADDRESSER: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
P.C.
STREET: 1755 S. Jefferson Davis Highway, Suite 400
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/810,501
FILING DATE: 12-Mar-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/301,435
FILING DATE: <Unknown>
APPLICATION NUMBER: 08/131,625
FILING DATE: 05-OCT-1993

ATTORNEY/AGENT INFORMATION:

NAME: Lavalleye, Jean-Paul M.P.
REGISTRATION NUMBER: 31,451
REFERENCE/DOCKET NUMBER: 4625-021-55X CIP

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220
TELEX: 248855 OPAT UR

INFORMATION FOR SEQ ID NO: 67:

SEQUENCE CHARACTERISTICS:

LENGTH: 256 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 67:

US-09-810-501-67

Query Match 5.9%; Score 77; DB 5; Length 256;
Best Local Similarity 18.8%; Pred. No. 2.7;
Matches 48; Conservative 33; Mismatches 74; Indels 100; Gaps 12;

QY 46 WVQKYEITPLLI-----CDTKFDIRQWFLVTDW-----NPLTIWFYKESYLR 88
DB 18 WMLSRSSWCLLISLYFWPFCCLASPSQVGMWSPASDFAPRYSVRALPFTLSNYSRYEA 77
QY 89 FSTQRFSLDKLDSAIHLCCNNAVQKY-LKNDVGRSPLLPAHNMWT-----STRFQYLOR 141
DB 78 FLSQ-----CQVDIPTWGTGKPLG---MLMHHKVSTLIDEMVSRMYRIMEK 121
QY 142 QGRGAVGMSVIVP-----SMKKAIAHAMKVAQDHVE-----PRKNSFELYGADFV 186
DB 122 AGQ-AAWKQVSEATLSRISLSDVVAHFQHLAAIEAETCKYLASRLPMLHLRMTGSNVT 180
QY 187 LG-----RDRPWLIEINSSPTMHPSTPVTVAQLCAQVQEDTIKV 225
DB 181 IVYNSTLNQVFAIFPTGSRPKLHDFQWLIAVHSS-----IFS 219
QY 226 AVDRSCDIGNFELLW 240
DB 220 SVAASCTL--FVVLW 232

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2006, 12:55:16 ; Search time 34.8606 Seconds
(without alignments)
2900.544 Million cell updates/sec

Title: US-10-635-977-2_COPY_133_374
Perfect score: 1299
Sequence: 1 IDGLNIIWIKPAKSRGRD.....IKVAVDRSCDIGNFELLWRQ 242

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA Main:
1: /cgn2_6/prodata/1/pubpaa/US07_PUBCOMB.pap.*
2: /cgn2_6/prodata/1/pubpaa/US08_PUBCOMB.pap.*
3: /cgn2_6/prodata/1/pubpaa/US09_PUBCOMB.pap.*
4: /cgn2_6/prodata/1/pubpaa/US10A_PUBCOMB.pap.*
5: /cgn2_6/prodata/1/pubpaa/US10B_PUBCOMB.pap.*
6: /cgn2_6/prodata/1/pubpaa/US11_PUBCOMB.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1299	100.0	242	4	US-10-615-659-4
2	1299	100.0	242	4	US-10-635-977-4
3	1299	100.0	541	4	US-10-615-659-2
4	1299	100.0	541	4	US-10-615-659-13
5	1299	100.0	541	4	US-10-635-977-2
6	1299	100.0	541	4	US-10-635-977-13
7	1247	96.0	293	4	US-10-615-659-14
8	1247	96.0	293	4	US-10-635-977-14
9	850	65.4	352	4	US-10-615-659-7
10	850	65.4	352	4	US-10-635-977-7
11	850	65.4	352	5	US-10-756-149-5267
12	850	65.4	362	3	US-09-925-298-617
13	850	65.4	362	4	US-10-102-806-617
14	850	65.4	488	4	US-10-275-595A-5
15	847	65.4	326	4	US-10-108-260A-4209
16	846	65.1	292	4	US-10-615-659-5
17	846	65.1	292	4	US-10-635-977-5
18	628.5	48.4	992	6	US-11-097-143-23727
19	499.5	38.5	719	6	US-11-097-143-23415
20	337	25.9	566	5	US-10-450-763-36330
21	324	24.9	496	6	US-11-097-143-19014
22	319	24.6	330	4	US-10-615-659-26
23	319	24.6	330	4	US-10-635-977-26
24	301	23.2	268	4	US-10-424-599-205823
25	297.5	22.9	827	6	US-11-097-143-9312
26	296	22.8	1226	5	US-10-756-149-5184
27	295.5	22.7	423	4	US-10-615-659-8

28	295.5	22.7	423	4	US-10-635-977-8	Sequence 8, Appli
29	284.5	21.9	989	6	US-11-097-143-7680	Sequence 7680, Ap
30	276	21.2	487	6	US-11-097-143-4080	Sequence 4080, Ap
31	274	21.1	49	4	US-10-615-659-21	Sequence 21, Appl
32	274	21.1	49	4	US-10-615-659-22	Sequence 22, Appl
33	274	21.1	49	4	US-10-635-977-21	Sequence 21, Appl
34	274	21.1	49	4	US-10-635-977-22	Sequence 22, Appl
35	271	20.9	439	4	US-10-104-047-3658	Sequence 3658, Ap
36	270	20.8	524	3	US-09-864-761-38213	Sequence 38213, A
37	270	20.8	592	4	US-10-104-047-3371	Sequence 3371, Ap
38	270	20.8	917	6	US-11-097-143-21303	Sequence 21303, A
39	258	19.9	92	4	US-10-424-599-262294	Sequence 262294, A
40	247.5	19.1	379	4	US-10-615-659-6	Sequence 6, Appli
41	247.5	19.1	379	4	US-10-635-977-6	Sequence 118, App
42	244.5	18.8	377	4	US-10-210-130-118	Sequence 1, Appli
43	244.5	18.8	377	4	US-10-250-613-1	Sequence 2495, Ap
44	205.5	15.8	553	4	US-10-108-260A-2495	Sequence 20, Appl
45	205.5	15.8	887	4	US-10-473-574-20	

ALIGNMENTS

RESULT 1
US-10-615-659-4
; Sequence 4, Application US/10615659
; Publication No. US20040157234A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; CURRENT FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4
; LENGTH: 242
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-615-659-4

Query Match	100.0%;	Score 1299;	DB 4;	Length 242;
Best Local Similarity	100.0%;	Pred. No. 3.2e-129;		
Matches 242;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	IDGLNIIWIKPAKSRGRDIVCMRVRVEILELAADHPLSRDNKWWVKYIETPLLICD	60	
Db	1	IDGLNIIWIKPAKSRGRDIVCMRVRVEILELAADHPLSRDNKWWVKYIETPLLICD	60	
Qy	61	TKFDIRQWFLVTDNPLTIWFKESYLRFTSRFSLDKLDSAIHLNNAVKYIKNDVGR	120	
Db	61	TKFDIRQWFLVTDNPLTIWFKESYLRFTSRFSLDKLDSAIHLNNAVKYIKNDVGR	120	
Qy	121	SPLLPAHNWNTSTRFOEYLORGRGAVGVSIVYPSMKKAIAHAMKVAODHVPKNSFEL	180	
Db	121	SPLLPAHNWNTSTRFOEYLORGRGAVGVSIVYPSMKKAIAHAMKVAODHVPKNSFEL	180	
Qy	181	YGADPVLGRDPRPWLIEINSSPTMHPSTFVTAQLCAQVQEDTIKVAVDKSCDIGNFELW	240	
Db	181	YGADPVLGRDPRPWLIEINSSPTMHPSTFVTAQLCAQVQEDTIKVAVDKSCDIGNFELW	240	
Qy	241	RQ 242		
Db	241	RQ 242		

RESULT 2
US-10-635-977-4
; Sequence 4, Application US/10635977
; Publication No. US20040171131A1

```

; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635,977
; CURRENT FILING DATE: 2003-08-07
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S. 10/615,659
; PRIOR FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4
; LENGTH: 242
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-635-977-4

```

```

Query Match      100.0%; Score 1299; DB 4; Length 242;
Best Local Similarity 100.0%; Pred. No. 3.2e-129;
Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IDGLRNIIWKPAKSRGRDIVCMDRVVEEILELAADHPILSRDNKVVQKIETPLICD 60
Db 1 IDGLRNIIWKPAKSRGRDIVCMDRVVEEILELAADHPILSRDNKVVQKIETPLICD 60

Qy 61 TKFDIRQWFLVTDWNPITWIFKESYLRFSTQRFSLDKLDSAIHLNNAVQKYLKNDVGR 120
Db 61 TKFDIRQWFLVTDWNPITWIFKESYLRFSTQRFSLDKLDSAIHLNNAVQKYLKNDVGR 120

Qy 121 SPLLPAHNMTSTRFQYLRQGRGAVWGSVIYPSMKKAIAHAMKVAQDHVEPRKNSFEL 180
Db 121 SPLLPAHNMTSTRFQYLRQGRGAVWGSVIYPSMKKAIAHAMKVAQDHVEPRKNSFEL 180

Qy 181 YGADFVLGRDPRPWLIEINSSPTMHPSTPVTQALCAQVQEDTIKVAVDKSCDIGNFELLW 240
Db 181 YGADFVLGRDPRPWLIEINSSPTMHPSTPVTQALCAQVQEDTIKVAVDKSCDIGNFELLW 240

Qy 241 RQ 242
Db 241 RQ 242

```

```

RESULT 3
US-10-615-659-2
; Sequence 2, Application US/10615659
; Publication No. US20040157234A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; CURRENT FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 541
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-615-659-2

```

```

Query Match      100.0%; Score 1299; DB 4; Length 541;
Best Local Similarity 100.0%; Pred. No. 9.5e-129;
Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IDGLRNIIWKPAKSRGRDIVCMDRVVEEILELAADHPILSRDNKVVQKIETPLICD 60
Db 133 IDGLRNIIWKPAKSRGRDIVCMDRVVEEILELAADHPILSRDNKVVQKIETPLICD 192

```

```

Qy 61 TKFDIRQWFLVTDWNPITWIFKESYLRFSTQRFSLDKLDSAIHLNNAVQKYLKNDVGR 120
Db 193 TKFDIRQWFLVTDWNPITWIFKESYLRFSTQRFSLDKLDSAIHLNNAVQKYLKNDVGR 252

Qy 121 SPLLPAHNMTSTRFQYLRQGRGAVWGSVIYPSMKKAIAHAMKVAQDHVEPRKNSFEL 180
Db 253 SPLLPAHNMTSTRFQYLRQGRGAVWGSVIYPSMKKAIAHAMKVAQDHVEPRKNSFEL 312

Qy 181 YGADFVLGRDPRPWLIEINSSPTMHPSTPVTQALCAQVQEDTIKVAVDKSCDIGNFELLW 240
Db 313 YGADFVLGRDPRPWLIEINSSPTMHPSTPVTQALCAQVQEDTIKVAVDKSCDIGNFELLW 372

Qy 241 RQ 242
Db 373 RQ 374

```

```

RESULT 4
US-10-615-659-13
; Sequence 13, Application US/10615659
; Publication No. US20040157234A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; CURRENT FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 13
; LENGTH: 541
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-615-659-13

```

```

Query Match      100.0%; Score 1299; DB 4; Length 541;
Best Local Similarity 100.0%; Pred. No. 9.5e-129;
Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IDGLRNIIWKPAKSRGRDIVCMDRVVEEILELAADHPILSRDNKVVQKIETPLICD 60
Db 133 IDGLRNIIWKPAKSRGRDIVCMDRVVEEILELAADHPILSRDNKVVQKIETPLICD 192

Qy 61 TKFDIRQWFLVTDWNPITWIFKESYLRFSTQRFSLDKLDSAIHLNNAVQKYLKNDVGR 120
Db 193 TKFDIRQWFLVTDWNPITWIFKESYLRFSTQRFSLDKLDSAIHLNNAVQKYLKNDVGR 252

Qy 121 SPLLPAHNMTSTRFQYLRQGRGAVWGSVIYPSMKKAIAHAMKVAQDHVEPRKNSFEL 180
Db 253 SPLLPAHNMTSTRFQYLRQGRGAVWGSVIYPSMKKAIAHAMKVAQDHVEPRKNSFEL 312

Qy 181 YGADFVLGRDPRPWLIEINSSPTMHPSTPVTQALCAQVQEDTIKVAVDKSCDIGNFELLW 240
Db 313 YGADFVLGRDPRPWLIEINSSPTMHPSTPVTQALCAQVQEDTIKVAVDKSCDIGNFELLW 372

Qy 241 RQ 242
Db 373 RQ 374

```

```

RESULT 5
US-10-635-977-2
; Sequence 2, Application US/10635977
; Publication No. US20040171131A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635,977

```


; CURRENT FILING DATE: 2003-08-07
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S. 10/615,659
; PRIOR FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 541
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-635-977-2

Query Match 100.0%; Score 1299; DB 4; Length 541;
Best Local Similarity 100.0%; Pred. No. 9.5e-129;
Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 IDGLNIIWIKPAKSRGRDIVCMRVEIELEAAADHPLSRDNKVVQKYEIETPLLCD 60
DB 133 IDGLNIIWIKPAKSRGRDIVCMRVEIELEAAADHPLSRDNKVVQKYEIETPLLCD 192
QY 61 TKFDIRQWFLVTDNPLTIWFYKESYLRFSTQRFSLDKLDSAIHLCLNNVQKYLKNDVGR 120
DB 193 TKFDIRQWFLVTDNPLTIWFYKESYLRFSTQRFSLDKLDSAIHLCLNNVQKYLKNDVGR 252
QY 121 SPILLPAHNNMTSTRFOEYLQRCRGAVGWSVIYPSMKKAIHAAMKVAQDHVPRKNSFEL 180
DB 253 SPILLPAHNNMTSTRFOEYLQRCRGAVGWSVIYPSMKKAIHAAMKVAQDHVPRKNSFEL 312
QY 181 YGADFVLGRDPRPWLIEINSSPTMHPSTPTVTAQLCAQVOEDTIKVAVDNRSCDIGNFELLW 240
DB 313 YGADFVLGRDPRPWLIEINSSPTMHPSTPTVTAQLCAQVOEDTIKVAVDNRSCDIGNFELLW 372
QY 241 RQ 242
DB 373 RQ 374

RESULT 6
US-10-635-977-13
; Sequence 13, Application US/10635977
; Publication No. US20040171131A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635,977
; PRIOR FILING DATE: 2003-08-07
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S. 10/615,659
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 13
; LENGTH: 541
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-635-977-13

Query Match 100.0%; Score 1299; DB 4; Length 541;
Best Local Similarity 100.0%; Pred. No. 9.5e-129;
Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 IDGLNIIWIKPAKSRGRDIVCMRVEIELEAAADHPLSRDNKVVQKYEIETPLLCD 60
DB 133 IDGLNIIWIKPAKSRGRDIVCMRVEIELEAAADHPLSRDNKVVQKYEIETPLLCD 192
QY 61 TKFDIRQWFLVTDNPLTIWFYKESYLRFSTQRFSLDKLDSAIHLCLNNVQKYLKNDVGR 120
DB 193 TKFDIRQWFLVTDNPLTIWFYKESYLRFSTQRFSLDKLDSAIHLCLNNVQKYLKNDVGR 252

QY 121 SPILLPAHNNMTSTRFOEYLQRCRGAVGWSVIYPSMKKAIHAAMKVAQDHVPRKNSFEL 180
DB 253 SPILLPAHNNMTSTRFOEYLQRCRGAVGWSVIYPSMKKAIHAAMKVAQDHVPRKNSFEL 312
QY 181 YGADFVLGRDPRPWLIEINSSPTMHPSTPTVTAQLCAQVOEDTIKVAVDNRSCDIGNFELLW 240
DB 313 YGADFVLGRDPRPWLIEINSSPTMHPSTPTVTAQLCAQVOEDTIKVAVDNRSCDIGNFELLW 372
QY 241 RQ 242
DB 373 RQ 374

RESULT 7
US-10-615-659-14
; Sequence 14, Application US/10615659
; Publication No. US20040157234A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; PRIOR FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 14
; LENGTH: 293
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-615-659-14

Query Match 96.0%; Score 1247; DB 4; Length 293;
Best Local Similarity 100.0%; Pred. No. 1.4e-123;
Matches 233; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IDGLNIIWIKPAKSRGRDIVCMRVEIELEAAADHPLSRDNKVVQKYEIETPLLCD 60
DB 61 IDGLNIIWIKPAKSRGRDIVCMRVEIELEAAADHPLSRDNKVVQKYEIETPLLCD 120
QY 61 TKFDIRQWFLVTDNPLTIWFYKESYLRFSTQRFSLDKLDSAIHLCLNNVQKYLKNDVGR 120
DB 121 TKFDIRQWFLVTDNPLTIWFYKESYLRFSTQRFSLDKLDSAIHLCLNNVQKYLKNDVGR 180
QY 121 SPILLPAHNNMTSTRFOEYLQRCRGAVGWSVIYPSMKKAIHAAMKVAQDHVPRKNSFEL 180
DB 181 SPILLPAHNNMTSTRFOEYLQRCRGAVGWSVIYPSMKKAIHAAMKVAQDHVPRKNSFEL 240
QY 181 YGADFVLGRDPRPWLIEINSSPTMHPSTPTVTAQLCAQVOEDTIKVAVDNRSCDI 233
DB 241 YGADFVLGRDPRPWLIEINSSPTMHPSTPTVTAQLCAQVOEDTIKVAVDNRSCDI 293

RESULT 8
US-10-635-977-14
; Sequence 14, Application US/10635977
; Publication No. US20040171131A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635,977
; PRIOR FILING DATE: 2003-08-07
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S. 10/615,659
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 14


```

; PRIOR APPLICATION NUMBER: US 60/201,960
; PRIOR FILING DATE: 2000-05-05
; PRIOR APPLICATION NUMBER: US 60/202,729
; PRIOR FILING DATE: 2000-05-08
; PRIOR APPLICATION NUMBER: US 60/209,705
; PRIOR FILING DATE: 2000-06-05
; PRIOR APPLICATION NUMBER: US 60/210,149
; PRIOR FILING DATE: 2000-06-07
; PRIOR APPLICATION NUMBER: US 60/213,215
; PRIOR FILING DATE: 2000-06-21
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: PERL Program
; SEQ ID NO 5
; LENGTH: 488
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No: 2156553CD1
US-10-275-595A-5

Query Match      65.4%; Score 850; DB 4; Length 488;
Best Local Similarity 61.8%; Pred. No. 4.1e-81;
Matches 152; Conservative 41; Mismatches 49; Indels 4; Gaps 1;

QY 1 IDGLRNIIWIKPAKSRGRDIVCMRVVEEILEAAADHPLSRDNKVVQKYIETPLLCD 60
DB 61 MEGDRNIWIKPGAKSRGRGIMCWDHLEMLKLVNGNPVVMKDGKVVQKYIERPLIFG 120
QY 61 TKFDIROWFLVTDWNLTIWTFYKESYLRFSSTQRFSLKQLDNSVHLCNNSIQKLENSCHR 120
DB 121 TKFDLRQWFLVTDWNLTIWTFYKESYLRFSSTQRFSLKQLDNSVHLCNNSIQKLENSCHR 180
QY 121 SPLLPAHNNMTSTRFOEYLORQGRGAVWGSVIYPSMKKAIAHAMKVAQDHVEPRKNSFEL 180
DB 181 HPLLPPDNMWSQRFOAHLQEMGAPNAWSTIIIVPGMKDAVIHALQTSQDTVQCRKASFEL 240
QY 181 YGADFVLGRDPRPWLIEINSSPTMHPSTPVTAQLCAQVQEDTIKVAV----DRSCDIGNF 236
DB 241 YGADFVFGVDFOFWLIEINASPTMAPSTAVTARLCAGVQADTLRVVIDRMLDRNCDTGAF 300
QY 237 ELLWRQ 242
DB 301 ELIYKQ 306

Search completed: April 4, 2006, 12:58:00
Job time : 34.8606 secs
```

```

; PRIOR APPLICATION NUMBER: US 60/201,960
; PRIOR FILING DATE: 2000-05-05
; PRIOR APPLICATION NUMBER: US 60/202,729
; PRIOR FILING DATE: 2000-05-08
; PRIOR APPLICATION NUMBER: US 60/209,705
; PRIOR FILING DATE: 2000-06-05
; PRIOR APPLICATION NUMBER: US 60/210,149
; PRIOR FILING DATE: 2000-06-07
; PRIOR APPLICATION NUMBER: US 60/213,215
; PRIOR FILING DATE: 2000-06-21
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: PERL Program
; SEQ ID NO 5
; LENGTH: 488
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No: 2156553CD1
US-10-275-595A-5

Query Match      65.4%; Score 850; DB 4; Length 488;
Best Local Similarity 61.8%; Pred. No. 4.1e-81;
Matches 152; Conservative 41; Mismatches 49; Indels 4; Gaps 1;

QY 1 IDGLRNIIWIKPAKSRGRDIVCMRVVEEILEAAADHPLSRDNKVVQKYIETPLLCD 60
DB 61 MEGDRNIWIKPGAKSRGRGIMCWDHLEMLKLVNGNPVVMKDGKVVQKYIERPLIFG 120
QY 61 TKFDIROWFLVTDWNLTIWTFYKESYLRFSSTQRFSLKQLDNSVHLCNNSIQKLENSCHR 120
DB 121 TKFDLRQWFLVTDWNLTIWTFYKESYLRFSSTQRFSLKQLDNSVHLCNNSIQKLENSCHR 180
QY 121 SPLLPAHNNMTSTRFOEYLORQGRGAVWGSVIYPSMKKAIAHAMKVAQDHVEPRKNSFEL 180
DB 181 HPLLPPDNMWSQRFOAHLQEMGAPNAWSTIIIVPGMKDAVIHALQTSQDTVQCRKASFEL 240
QY 181 YGADFVLGRDPRPWLIEINSSPTMHPSTPVTAQLCAQVQEDTIKVAV----DRSCDIGNF 236
DB 241 YGADFVFGVDFOFWLIEINASPTMAPSTAVTARLCAGVQADTLRVVIDRMLDRNCDTGAF 300
QY 237 ELLWRQ 242
DB 301 ELIYKQ 306
```

```

RESULT 15
US-10-108-260A-4209
; Sequence 4209, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20040005560A1el full length cDNA
; FILE REFERENCE: HI-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4209
; LENGTH: 326
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-108-260A-4209

Query Match      65.2%; Score 847; DB 4; Length 326;
Best Local Similarity 61.8%; Pred. No. 4.9e-81;
Matches 152; Conservative 41; Mismatches 49; Indels 4; Gaps 1;

QY 1 IDGLRNIIWIKPAKSRGRDIVCMRVVEEILEAAADHPLSRDNKVVQKYIETPLLCD 60
DB 61 MEGDRNIWIKPGAKSRGRGIMCWDHLEMLKLVNGNPVVMKDGKVVQKYIERPLIFG 120
QY 61 TKFDIROWFLVTDWNLTIWTFYKESYLRFSSTQRFSLKQLDNSVHLCNNSIQKLENSCHR 120
```

```

Query Match      65.2%; Score 847; DB 4; Length 326;
Best Local Similarity 61.8%; Pred. No. 4.9e-81;
Matches 152; Conservative 41; Mismatches 49; Indels 4; Gaps 1;

QY 1 IDGLRNIIWIKPAKSRGRDIVCMRVVEEILEAAADHPLSRDNKVVQKYIETPLLCD 60
DB 61 MEGDRNIWIKPGAKSRGRGIMCWDHLEMLKLVNGNPVVMKDGKVVQKYIERPLIFG 120
QY 61 TKFDIROWFLVTDWNLTIWTFYKESYLRFSSTQRFSLKQLDNSVHLCNNSIQKLENSCHR 120
```

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2006, 12:46:06 ; Search time 9.4461 Seconds
(without alignments)
2118.072 Million cell updates/sec

Title: US-10-635-977-2_COPY_133_374

Perfect score: 1299

Sequence: 1 IDGLRNWIKPAKSRGRD.....IKVAVDRSCDIGNFELLMRQ 242

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/prodata/1/1aa/5_COMB.pep.*
- 2: /cgn2_6/prodata/1/1aa/6_COMB.pep.*
- 3: /cgn2_6/prodata/1/1aa/H_COMB.pep.*
- 4: /cgn2_6/prodata/1/1aa/PCITUS_COMB.pep.*
- 5: /cgn2_6/prodata/1/1aa/RE_COMB.pep.*
- 6: /cgn2_6/prodata/1/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	271	20.9	439	2	US-10-104-047-3658
2	270	20.8	592	2	US-10-104-047-3371
3	212	16.3	561	2	US-09-270-767-46703
4	177	13.6	404	2	US-10-104-047-3621
5	153.5	11.8	260	2	US-09-270-767-62411
6	153.5	11.8	507	2	US-09-270-767-46787
7	149	11.5	259	2	US-09-248-796A-14603
8	144	11.1	753	2	US-09-538-092-31
9	126	9.7	305	2	US-09-270-767-46577
10	102	7.9	55	2	US-09-270-767-62317
11	91.5	7.0	816	2	US-09-248-796A-20939
12	87.5	6.7	607	2	US-09-489-039A-12140
13	83.5	6.4	268	1	US-07-857-224B-33
14	83	6.4	732	2	US-08-671-757A-7
15	83	6.4	732	2	US-08-671-757A-8
16	83	6.4	732	2	US-09-015-078-7
17	83	6.4	732	2	US-09-015-078-8
18	83	6.4	732	2	US-10-238-977A-7
19	83	6.4	732	2	US-10-238-977A-8
20	80	6.2	243	2	US-09-902-540-15153
21	80	6.2	1577	1	US-08-793-824-2
22	79.5	6.1	433	2	US-09-344-882-26
23	79.5	6.1	433	2	US-10-293-865-26
24	79	6.1	513	2	US-09-351-229-4
25	78.5	6.0	477	2	US-09-075-375F-4
26	78	6.0	256	1	US-08-799-464A-3
27	78	6.0	256	2	US-08-478-316-9

28	78	6.0	256	2	US-08-478-316-14	Sequence 14, Appl
29	78	6.0	256	2	US-09-113-750A-37	Sequence 37, Appl
30	78	6.0	256	2	US-09-019-793A-9	Sequence 9, Appl
31	78	6.0	256	2	US-09-019-793A-14	Sequence 14, Appl
32	78	6.0	256	2	US-09-601-328-9	Sequence 9, Appl
33	78	6.0	256	2	US-09-601-328-14	Sequence 14, Appl
34	78	6.0	256	4	PCT-US95-09927-3	Sequence 3, Appl
35	78	6.0	667	2	US-09-710-279-198	Sequence 198, App
36	77.5	6.0	320	2	US-09-605-703B-808	Sequence 808, App
37	77.5	6.0	320	2	US-09-605-703B-810	Sequence 810, App
38	77.5	6.0	965	2	US-09-252-991A-24038	Sequence 24038, A
39	77	5.9	256	2	US-08-478-316-8	Sequence 8, Appl
40	77	5.9	256	2	US-08-478-316-11	Sequence 11, Appl
41	77	5.9	256	2	US-09-019-793A-8	Sequence 8, Appl
42	77	5.9	256	2	US-09-019-793A-11	Sequence 11, Appl
43	77	5.9	256	2	US-08-301-435-67	Sequence 67, Appl
44	77	5.9	256	2	US-09-601-328-8	Sequence 8, Appl
45	77	5.9	256	2	US-09-601-328-11	Sequence 11, Appl

ALIGNMENTS

RESULT 1

US-10-104-047-3658
; Sequence 3658, Application US/10104047
; Patent No. 6943241
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: NO. 6943241el full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 3658
; LENGTH: 439
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-104-047-3658

Query Match 20.9%; Score 271; DB 2; Length 439;
Best Local Similarity 29.8%; Pred. No. 3.8e-22;
Matches 75; Conservative 45; Mismatches 86; Indels 46; Gaps 9;

QY	8	WIIPAAKSRGRDIVCMRVEIIELEAAADHPLSRDNK-----WVQKYIETPLLCD 60
DB	164	WIMKPVARSQKGIFLFRLLKDIVDW-RKTRSSDDQKDDIPVENYVAQRYIENPYLGG 222
QY	61	TKEDIQWFLVTDNPLTIWFKESYLFSTQFSLDKLDSAIHLCLNNAVQKYLKNDVCR 120
DB	223	RKFDLRVLYVMS-----VFAECLLWSGHR-----QDVHLTNVAVQK----- 260
QY	121	SPLPAHN-----MWTSTRFOEYL--QRQGRGAVGWSVIYPSMKKAIAHAMKVAQDHVEPR 174
DB	261	--TFSDYHPKKGCKWTLQRFQYLAASKHGPEAV--ETLFRIDNITFVKLSQVQKVIISD 316
QY	175	KNSFELYGADFLVGRDFRPWLIEINSSFTMPHTPVTVAQLCAQVOEDTIKVA-----VD 228
DB	317	KHCFELYGYDILIDQLKPLLEVNASPSLTASSQEDVELKTCLEDTLHVVDMEARLTG 376
QY	229	RSCDIGNFELMW 240
DB	377	REKRVGGFDLMW 388

RESULT 2

US-10-104-047-3371
; Sequence 3371, Application US/10104047
; Patent No. 6943241
; GENERAL INFORMATION:

[illegible]

```

QY 132 STRFQEYLQRFGRGAVGWGSGVIYPSMKKAIHAAMKVAQ--DHVEPRKNSFELYGADFVLGR 189
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 557 IEFNEFYMSDRKGF---SSFKLIDNGYDAIKTKETVTKRKSRKSTSAFNKDI---S 610
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 190 DFRPWLIEINSSPTMHP 207
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 611 DFRQGLRRENNGESFPPT 628
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 12
US-09-489-039A-12140
; Sequence 12140, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 12140
; LENGTH: 607
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-12140

Query Match 6.7%; Score 87.5; DB 2; Length 607;
Best Local Similarity 25.2%; Pred. No. 0.61;
Matches 52; Conservative 23; Mismatches 98; Indels 33; Gaps 11;

QY 8 WI----IKPAKSRGRDVCMDRVEEILELAAADHPLSRDNKMWVQKYEIETPLLICDTKFD 64
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 175 WVSLLTDPVSGMARDVVLSDAMGYLHFISGIP--TGTNWL----YSTPYKMATPPLS 229
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 65 -IROQFLVTDWN--PLTIWFKESYLRFSTQRFSLDKLDSAIHLCNNAVQKYLKNDVGRS 121
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 230 VINQWLALDNGSLPAFIAGLAPRHPOVETMHQSL-----LALVADSRPWPQMTSG 281
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 122 PLLPAHNWTSF--RFQEYLQRFGRGAVGWGSGVIYPSMKKAIHAAMKVAQDH--V 171
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 282 SLRFECE--WSNDIGALREILORTGMLENSANIVLPEDVVSFAKKSRKSPAAKGVYDRQLV 339
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 172 E--PRKNSFELYGADFVLGRDPRFWL 195
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 340 EGVKRFQAMQGLGADGVIGQSTRDWL 365
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 13
US-07-857-224B-33
; Sequence 33, Application US/07857224B
; Patent No. 5958784
; GENERAL INFORMATION:
; APPLICANT: Benner, Steven A.
; TITLE OF INVENTION: Predicting Folded Structures of Proteins
; NUMBER OF SEQUENCES: 114
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Steven A. Benner
; STREET: Hadlaubstrasse 151
; CITY: Zurich
; STATE: none
; COUNTRY: Switzerland
; ZIP: (note: this is an international post code) CH-8092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch diskette, 1.4 Mb storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 7.0
; SOFTWARE: Microsoft Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/857,224B
; FILING DATE: 03/25/92

```

CLASSIFICATION: 436
PRIOR APPLICATION DATA: none
TELECOMMUNICATION INFORMATION:
TELEPHONE: (International) 41 1 632 2830
TELEFAX: (International) 41 1 262 2437
TELEX: none
INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:
LENGTH: 268
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
DESCRIPTION: protein
ORIGINAL SOURCE:
ORGANISM: Saccharomyces cerevisiae
FEATURE: Protein kinase; Table 8 Column 37
PUBLICATION INFORMATION:
AUTHORS:
AUTHORS: Hanks, S. K.
AUTHORS: Quinn, A. M.
AUTHORS: Hunter, T.
TITLE: The protein kinase family
JOURNAL: Science
VOLUME: 241
PAGES: 42-52
DATE: 1988
US-07-857-224B-33

Query Match 6.4%; Score 83.5; DB 1; Length 268;
Best Local Similarity 21.4%; Pred. No. 0.52; Indels 65; Gaps 11;
Matches 54; Conservative 33; Mismatches 100; Indels 65; Gaps 11;
QY 10 IKPAKSRGRDIVCMRVEEILELAAADHP-----LSRDNKWVQKYEITPL---- 56
DB 31 IKEITSEFKGLDMSAIREVKYLOEQHPNVEILIDIFMAYDNLNLVLEFLPTLEVI 90
QY 57 ---LICTDKFDIRQNFVLT-----DWNELTTWFKESYLRSTQFSLDK 98
DB 91 KDKSLFTPADIKAWMLMTLRGVYCHRNFIHLRDLKPNLLFSPDGGQIKVAD--FGLR 148
QY 99 LDSAIH--LCNNAVOKLKND---VGRSPLLPAHNMWT--STRFOEYLOR---QCRGAVW 148
DB 149 AIPAPHEILTSNVVTRWRAPELLFGAKHYTSALDINSVGVIPLAEMLRIPYLPQNDV- 207
QY 149 GSVIYPSMKKAIAHAMKVAQDHVEPRKNSF---ELYGADFVLGRDFRPWLEINSPTWH 205
DB 208 -----DQMEVTFRALGTPTRDWDPEVSSFWTASEYALDFMCG-----MLTWN 249
QY 206 PSTPVTQAOLCAQ 217
DB 250 PQRWTAVQCLE 261

RESULT 14
US-08-671-757A-7
Sequence 7, Application US/08671757A
Patent No. 6476213
GENERAL INFORMATION:
APPLICANT: Suerbaum, Sebastian
TITLE OF INVENTION: Cloning and Characterization Production
of Aflagellate Strains
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
Dunner
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/671,757A
FILING DATE: 16-Aug-2001
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/671,757A
FILING DATE: 16-Aug-2001
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J.
REFERENCE/DOCKET NUMBER: 02356.0073-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 732 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-08-671-757A-7

Query Match 6.4%; Score 83; DB 2; Length 732;
Best Local Similarity 24.5%; Pred. No. 2.6;
Matches 51; Conservative 31; Mismatches 74; Indels 52; Gaps 12;
QY 15 KSRGRDIVCMRVEEILELAAADHP-LSRDNKWWVQKYEITPL-LICTDKFDIRQNF--- 69
DB 531 KKYAEDFTIKDEVKSLLELRADYPTIVEESKKIPTGAIRSVLQALLHEKIPKMLTIL 590
QY 70 -LVTDWNL-----TWFKY-----SYLRFST--QRFSLDKLSA 102
DB 591 ETITDIALVQNDVNILTEQVRLSRVITNAFSEGDGLKELTSTDEQFLLNKLR-- 648
QY 103 IHLCNNAVOKLKNDVGR-SPLLPAHNMWTSTRFOEYLORQCRGAVMSVIYPSMKKAIA 161
DB 649 ---ENGTSKSLLLNVGELQKLEA---VSEEMKVLQKGIAPVI--LIVEPNLRKALS 698
QY 162 HAMKVAQ-----DHVEPRKNS-PELYG 182
DB 699 NQMEQARIDIVLSHAELDPNSFEALG 726

RESULT 15
US-08-671-757A-8
Sequence 8, Application US/08671757A
Patent No. 6476213
GENERAL INFORMATION:
APPLICANT: Suerbaum, Sebastian
TITLE OF INVENTION: Cloning and Characterization Production
of Aflagellate Strains
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
Dunner
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/671,757A
FILING DATE: 16-Aug-2001
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146

```

; REFERENCE/DOCKET NUMBER: 02356.0073-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 732 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-08-671-757A-8

Query Match      6.4%; Score 83; DB 2; Length 732;
Best Local Similarity 24.5%; Pred. No. 2.6;
Matches 51; Conservative 31; Mismatches 74; Indels 52; Gaps 12;

QY 15 KSRGRDIVCMRVVEEILELAAADHP-LSRDNKWVQKYYETPL-LICDTKFDIRQWF--- 69
Db 531 KKYAEDFITKDEVKSLERLAKDPTIVEESKKIPTGAIRSVLQALLHEKIPKIDMLTIL 590
QY 70 -LVTDNPL-----TIFVKE----SYLRFST--ORFSLDKLDSA 102
Db 591 ETITDIAPLVQNDVNILTEQVRARLSRVITNAFKSEDCGRCLKPLTFSTDSEQFLNKLKLR-- 648
QY 103 IHLNNAVQKYLKNDVGR-SPLLPFAHNMWTSRFOEYLQQRGAVNGSVIYPSMKKAIA 161
Db 649 -----ENGTSKSLLNVLNGELQKLIEA---VSEEMKVLQKGIAPVI--LIVEPNLRKALS 698
QY 162 HAMKVAQ-----DHVEPRKNS-FELYG 182
Db 699 NOMEQARIDVIVLSHAELDPNSNFEALG 726
```

Search completed: April 4, 2006, 12:46:56
Job time : 9.4461 secs

GenCore version 5.1.7

Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2006, 12:43:06 ; Search time 36.6599 Seconds
(without alignments) 4657.352 Million cell updates/sec

Title: US-10-635-977-2 COPY 133 374

Perfect score: 1299
Sequence: 1 IDGLRNIWIIKPAKSRGRD.....IKVAVDRSCDIGNFELLWRO 242

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

```

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                  Maximum Match 10%
                  Listing first 45

```

```
Database : UniProt_05.80:*
1: uniprot_sprot:*
2: uniprot_trembl:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %		DB	ID	Description
		Match	Length			
1	1287	99.1	747	2	Q5JZ84_HUMAN	O5JZ84 homo sapien
2	1143	88.0	518	2	Q8C0V2_MOUSE	Q8C0V2 mus musculus
3	1143	88.0	781	2	Q8CON7_MOUSE	Q8CON7 mus musculus
4	850	65.4	352	1	TTLJ3_HUMAN	Q9Y4r7 homo sapien
5	850	65.4	352	2	Q6AWA3_HUMAN	Q6AWA3 homo sapien
6	850	65.4	434	2	Q4KMS8_HUMAN	Q4KMS8 homo sapien
7	850	65.4	434	2	Q8NDN8_HUMAN	Q8NDN8 homo sapien
8	850	65.4	744	2	Q9H876_HUMAN	Q9H876 homo sapien
9	838	64.5	704	2	Q8BV51_MOUSE	Q8BV51 mus musculus
10	834	64.2	266	2	Q922T0_MOUSE	Q922T0 mus musculus
11	799	61.5	261	2	Q922CT_BOVIN	Q922CT bos taurus
12	786	60.5	534	2	Q4RYO8_TETNG	Q4RYO8 tetraodon n
13	644.5	49.6	572	2	Q7OL56_ANOGA	Q7OL56 anopheles g
14	628.5	48.4	992	2	Q9VM91_DROME	Q9VM91 drosophila
15	521	40.1	501	2	Q7PMD3_ANOGA	Q7PMD3 anopheles g
16	499.5	38.5	719	2	Q9VM92_DROME	Q9VM92 drosophila
17	499.5	38.5	756	2	Q5BHY1_DROME	Q5BHY1 drosophila
18	468	36.0	281	2	Q5SHY1_DROME	Q5SHY1 drosophila
19	423	32.6	331	2	Q5TNZ9_ANOGA	Q5TNZ9 homo sapien
20	385	29.6	101	2	Q96GG8_HUMAN	Q96GG8 anopheles g
21	342	26.3	461	2	Q96GSG_HUMAN	Q96GSG homo sapien
22	342	26.3	461	2	Q641W7_RAT	Q641W7 rattus norv
23	340.5	26.2	464	2	Q9D570_MOUSE	Q9D570 mus musculus
24	328.5	25.3	523	2	Q5ATUI_DICDI	Q5ATUI dictyosteli
25	324	24.9	794	2	Q6BFH6_PARTE	Q6BFH6 paramesiciu
26	324	24.9	496	2	Q961I9_DROME	Q961I9 drosophila
27	320.5	24.7	1075	2	Q9VX74_DROME	Q9VX74 drosophila
28	313	24.1	341	2	Q7OTI3_GIALA	Q7OTI3 giardia lam
29	299	23.0	1339	2	Q5VX47_HUMAN	Q5VX47 homo sapien
30	297.5	22.9	827	2	Q5F498_CHICK	Q5F498 gallus gall
31	297.5	22.9	828	2	Q9VKL9_DROME	Q9VKL9 drosophila
			828	2	Q81GW4_DROME	Q81GW4 drosophila

RESULT 1

```

Q5JZ84 HUMAN
ID Q5JZ84 HUMAN PRELIMINARY; PRT; 747 AA.
AC Q5JZ84;
DT 10-MAY-2005 (TrEMBLrel. 30, Created)
DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE OTTHUMP0000028514 (Fragment).
DN ORFNames=RP3-355C18.2-002;
GN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
[1] _ NUCLEOTIDE SEQUENCE.
RA Cobley V.;
RL Submitted (MAY-2005) to the ENBL/GenBank/DBBJ databases.
DR EMBL; ALU022327; CAI42686.1; -; Genomic DNA.
DR GO; GO:0004835; F:tubulin-tyrosine ligase activity; IEA.
DR GO; GO:0006464; P:protein modification; IEA.
DR InterPro; IPR004344; TUB tyr lygase.

```

FI	NON	IER	I	I	I
SQ	SEQUENCE	747	AA:	83854	MW: F90948E159BBE589 CRC64:

Query Match 99.1%; Score 1287; DB 2; Length 747;
Best Local Similarity 98.4%; Pred. No. 9.5e-107;
Matches 242; Conservative 0; Mismatches 0; Indels

1	QY	IDGLRNIWIKPAAKSRGR----	DIVCMRVSEILEAAADHPLSRDNKWWVQKIETPL	56
335	Db	IDGLRNIWIKPAAKSRGRGESPD	IVCMRVSEILEAAADHPLSRDNKWWVQKIETPL	394
57	QY	LICDTPDIROWFLVTWNPLTT	TFWYKESYLRSTORFSLDKLSDSAIHLCNNAVQYKLN	116
395	Db	LICDTPDIROWFLVTWNPLTT	TFWYKESYLRSTORFSLDKLSDSAIHLCNNAVQYKLN	454
117	QY	DVGRSPLLPAINMTSTRFOEYL	QROGRGAVGWSVYPSMKKAIAHAMKVAQDHPBPRKN	176
455	Db	DVGRSPLLPAINMTSTRFOEYL	QROGRGAVGWSVYPSMKKAIAHAMKVAQDHPBPRKN	514
177	QY	SFELYGADFVLGRDFRFLWLT	INSSPTMHFSTPTVTAQLCAQVQEDTIKVAVDRCSDIGNF	236
515	Db	SFELYGADFVLGRDFRFLWLT	INSSPTMHFSTPTVTAQLCAQVQEDTIKVAVDRCSDIGNF	574
237	QY	ELLWRQ	242	
575	Db	ELLWRQ	580	

RESULT 2

Q8C0V2 MOUSE
ID Q8C0V2 MOUSE PRELIMINARY; PRT; 518 AA.

ALIGNMENTS

PRT; 518 AA.

AC Q8C0V2; NUCLEOTIDE SEQUENCE.
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Mus musculus adult male testis cDNA, RIKEN full-length enriched
 DE library, clone:4930524K07 product:hypothetical Tubulin-tyrosine ligase
 DE containing protein, full insert sequence.
 DE Name:1700019P01Rik;
 OS Mus musculus (Mouse)
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
 OC Muridea; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RN NUCLEOTIDE SEQUENCE.
 RP STRAIN=C57BL/6J; TISSUE=Testis;
 RC MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
 RA Carninci P., Hayashizaki Y.;
 RT "High-efficiency full-length cDNA cloning.";
 RL Meth. Enzymol. 303:19-44(1999).
 RN [2]
 RN NUCLEOTIDE SEQUENCE.
 RP STRAIN=C57BL/6J; TISSUE=Testis;
 RC MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staebli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Saeki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 RN [3]
 RN NUCLEOTIDE SEQUENCE.
 RP STRAIN=C57BL/6J; TISSUE=Testis;
 RC The FANTOM Consortium,
 RA The RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs.";
 RL Nature 420:563-573(2002).
 RN [4]
 RN NUCLEOTIDE SEQUENCE.
 RP STRAIN=C57BL/6J; TISSUE=Testis;
 RC MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
 RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to
 RT prepare full-length cDNA libraries for rapid discovery of new genes.";
 RL Genome Res. 10:1617-1630(2000).
 RN [5]
 RN NUCLEOTIDE SEQUENCE.
 RP STRAIN=C57BL/6J; TISSUE=Testis;
 RC MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
 RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P.,
 RA Konno H., Akiyama J., Nishi K., Kitsuunai T., Tashiro H., Itoh M.,
 RA Suni N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
 RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-format

RT sequencing pipeline with 384 multicapillary sequencer.";
 RL Genome Res. 10:1757-1771(2000).
 RN [6]
 RN NUCLEOTIDE SEQUENCE.
 RP STRAIN=C57BL/6J; TISSUE=Testis;
 RC Adachi J., Aizawa K., Akimura T., Hara A., Hashizume W.,
 RA Fukuda S., Furuno M., Hanagaki T., Hiraoka T., Hirozane T.,
 RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Kasukawa T.,
 RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
 RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
 RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
 RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
 RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
 RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
 RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
 RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK029745; BC26595.1; -; mRNA.
 DR Ensembl; ENSMUSG0000022388; Mus musculus.
 DR MGI; MGI:1922902; 1700019P01Rik
 DR GO; GO:0016874; F.ligase activity; IEA.
 DR GO; GO:0004835; F.tubulin-tyrosine ligase activity; IEA.
 DR GO; GO:0006464; P.protein modification; IEA.
 DR InterPro; IPR004344; Tub_tyr_ligase.
 DR Pfam; PF03133; TTL; 1.
 KW Hypothetical protein; Ligase.
 SQ SEQUENCE 518 AA; 59319 MW; 540C045659FAC0D6 CRC64;
 Query Match 88.0%; Score 1143; DB 2; Length 518;
 Best Local Similarity 85.0%; Pred. No. 5.4e-94;
 Matches 209; Conservative 17; Mismatches 16; Indels 4; Gaps 1;
 Qy 1 IDGRNIWIIPKPAKGRDIVCMWRVEETLELAADHPLSRDNKVVQVQKIETPLIICD 60
 Db 189 IDGRNIWIIPKPAKGRDIVCMWRVENILSVAADSTQTKONKVVQVQKIETPLIYD 248
 Qy 61 TKFDIQWFLVTDNPLTIWFYKESYLFSTQFSLDKLDSAIHLNNAVQKYLKNDVGR 120
 Db 249 TKFDIQWFLVTDNPLTIWFYKESYLFSTQFSLDKLDSAIHLNNSIQRLKNDKER 308
 Qy 121 SPLLPAAHNMWTSRFOEYLQQRGVAVGWSVIYPSMKKAIAHAMKVAQDHVEPRKNSFEL 180
 Db 309 SPLLPCHNMWTSRFOEYLQKRGRTGWSIIYPSMKRAVTNMRVAQDHVEARKNSFEL 368
 Qy 181 YGADFVLGRDPRFWLIEINSSPTMHSTPTVTAQLCAQVQEDTIKVAV----DRSCDIGNF 236
 Db 369 YGADFILGRDFKPLWIEINSSPTMHPSTPTVTAQLCAQVQEDTIKVAVVVDRLDRNCIDIGNF 428
 Qy 237 ELLWRQ 242
 Db 429 ELLWRQ 434
 RESULT 3
 Q8CON7_MOUSE PRELIMINARY; PRT; 781 AA.
 ID Q8CON7_MOUSE PRELIMINARY;
 AC Q8CON7;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Mus musculus adult male testis cDNA, RIKEN full-length enriched
 DE library, clone:4933401B17 product:hypothetical Tubulin-tyrosine ligase
 DE containing protein, full insert sequence.
 DE Name:1700019P01Rik;
 OS Mus musculus (Mouse).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
 OC Muridea; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RN NUCLEOTIDE SEQUENCE.
 RP STRAIN=C57BL/6J; TISSUE=Testis;
 RC MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
 RA Carninci P., Hayashizaki Y.;

[illegible]

Query Match	65.4%;	Score	850;	DB	2;	Length	352;
Best Local Similarity	61.8%;	Pred.	No. 7.6e-68;				
Matches	152;	Conservative	41;	Mismatches	49;	Indels	4;
Gaps	1;						
Qy	1	IDLRLNIWIKPAAKSRGRDIVCMDRVVEEILELAADHPLSRDNKWVQVQKVIETPLLICD	60				
Db	61	MEGRNIWIKVPGAKSRGRGIMCNDHLEMLKLVNGPNVVMKDGKVVQKVIIEPPLIFG	120				
Qy	61	TKKDIRQWFLVTDWNPITIVFYKESYRFTQPSFLDKLDSAIHLCNNVQVYKLVNDVGR	120				
Db	121	TKKDLRQWFLVTDWNPITVWVFYRDSYTRFSTQPSFLKNDLNSVHLCCNNSIQKILENSCHR	180				
Qy	121	SPLLPAHNMTSTRFQBYLQQRGAVGWSVYPSMKKAIHAHMKVAQDHVEPRKNSFEL	180				
Db	181	HPLLPPDNWSSQRFQAHQEMGAPNAWSTIIVPGMKDAVIHALQTSQDVTQCEKASFEL	240				
Qy	181	YGADFVLGRDFRFLWLEINSPTWHFSTPTVTAQLCAQVQEDTIKVAV-----DRSCDIGNF	236				
Db	241	YGADFVFGEDFQWLLIBINASTMAPSTAVTARLCAGVQADTLRWIDRLDRNCIDTGAF	300				
Qy	237	ELLWRQ	242				
Db	301	ELIYKO	306				

RESULT 7	Q8NDN8_HUMAN	Q8NDN8_HUMAN PRELIMINARY;	PRT;	434 AA.
ID	Q8NDN8	Q8NDN8		
AC	Q8NDN8	Q8NDN8		
DC	Q8NDN8	Q8NDN8		
DT	01-OCT-2002	(TrEMBLrel. 22, Created)		
DT	01-OCT-2002	(TrEMBLrel. 22, Last sequence update)		
DT	13-SEP-2005	(TrEMBLrel. 31, Last annotation update)		
DE	Hypothetical protein DKFZp566B0320.			
GN	Name=DKFZp566B0320;			
OS	Homo sapiens (Human)			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;			
OC	Homo			
OX	NCBI_TaxID=9606;			
OX	[1]			
RN	NUCLEOTIDE SEQUENCE.			
RP	TISSUE=PCR rescued clones;			
RP	TISSUE=Uterus;			
RC	Wambutt R., Heubner D., Mewes H.W., Gassenhuber J., Wiemann S.;			
RL	Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.			
RL	[2]			
RN	NUCLEOTIDE SEQUENCE.			
RP	TISSUE=PCR rescued clones;			
RP	MEDLINE=2238257; PubMed=12477932; DOI=10.1073/pnas.242603899;			
RX	Straussberg R.L., Fingold E.A., Grouse L.H., Derge J.G.,			
RA	Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,			
RA	Altshul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhac N.K.,			
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,			
RA	Diatchenko L., Marusina K., Farmer A.A., Casavant T.L., Schetz T.E.,			
RA	Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,			
RA	Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,			
RA	Raha S.C., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,			
RA	Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,			
RA	Richardson S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,			
RA	Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,			
RA	Fahey J., Helton E., Kettner M., Madan A., Young A.C., Gibbs R.A.,			
RA	Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,			
RA	Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,			
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,			
RA	Butterfield Y.S.N., Krzyzanski M.I., Skalska U., Smalus D.E.,			
RA	Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;			
RT	"Generation and initial analysis of more than 15,000 full-length human			
RT	and mouse cDNA sequences."			
RT	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).			
RT	[3]			
RN	NUCLEOTIDE SEQUENCE.			
RP	TISSUE=PCR rescued clones;			
RC	NIH MGC Project;			
RG	NIH submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.			

```

DR ENBL; ALB33399; CAD38794.1; -; mRNA.
DR ENBL; BC098298; AAH98298.1; -; mRNA.
DR GO; GO:0004935; F:ubulin-tyrosine ligase activity; IEA.
DR GO; GO:0006464; P:protein modification; IEA.
DR InterPro; IPR004344; Tub_tyr_ligase.
DR Pfam; PF03133; TTL; 1.
DR Hypothetical protein.
DR SQ SEQUENCE 434 AA; 49433 MW; 9E79BE6CA08651CA1 CRC64;

Query Match          65.4%; Score 850; DB 2; Length 434;
Best Local Similarity 61.8%; Pred. No. 9.7e-68;
Matches 152; Conservative 41; Mismatches 49; Indels 4; Gaps 1

QY 1 IDGLRNIWIKPAKSRGRDIVCMRVEEIELEAAADHPLSRDNKVVVQKVIETPLLICD 60
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 61 MEGDRNIWI VKPKAKSRGRGIMCWDHLEEMLKL VNGNPVMVKDGKVVVQKVIETPLLI 120

QY 61 TKEDIRQWFLVTDWNPDLTIWFYKSYLRFTSRFSLDKLSAIHL CNNAVOKYLNVDVR 120
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 121 TKEDLRQWFLVTDWNPDLTIWVFDSDYTRFSTQPFSLKNLONS VHLCNNSIQKHLNSCHR 180

QY 121 SPLLPAHNWMTSTRFOEYLRQCRGAVMGSVIVPSMKKATAHAMKVADHDVPEPKNSPEL 180
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 181 HPLLPPDNWSSSRFQAHLQEMGAPNAWSTII VPGMKDAVIHALQTSQDTVQCRKASPEL 240

QY 181 YGADFLVGLDRFRPWLIIENSSPTTHPSTPTVAQLCAQVQEDTIKAVV----DRSCDIGNF 236
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 241 YGADFVFGEDFPWLIIENASPTMAWSTAVTARLCAGVQADTLRVVIDRLDRNCDTGAF 300

QY 237 ELLWRQ 242
   |||||
Db 301 ELIIYKQ 306

```

RESULT	8
ID	Q9H876 HUMAN
AC	Q9H876 HUMAN PRELIMINARY; PRT; 744 AA.
DC	Q9H876;
DT	01-WAR-2001 (TRENBLrel. 16, Created)
DT	01-WAR-2001 (TRENBLrel. 16, Last sequence update)
DT	01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE	Hypothetical protein FLJ13898.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae;
OC	Homo.
OX	NCBI_TaxID=9606;
RN	[1]
RP	NUCLEOTIDE SEQUENCE.
RC	TISSUE=Thyroid gland;
RC	PubMed=14702039; DOI=10.1038/ngi285;
RA	Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA	Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA	Sekine M., Obayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
RA	Yamamoto J.-I., Saito K., Kawai Y., Isono Y., Nakamura Y.,
RA	Nagahori K., Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M.,
RA	Shiratori A., Sudoh H., Hosoiri T., Kaku Y., Kodaira H., Kondo H.,
RA	Sugawara M., Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E.,
RA	Omura Y., Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M.,
RA	Yamazaki M., Ninomiya K., Ishibashi T., Yamashita H., Murakawa K.,
RA	Fujimori K., Tanai H., Kimata M., Watanabe M., Hirooka S., Chiba Y.,
RA	Ishida S., Ono Y., Takiguchi S., Watanabe S., Yoshida M., Hotuta T.,
RA	Kusano J., Kanehori K., Takahashi-Fujii A., Hara H., Tanase T.-O.,
RA	Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K., Ariwa M.,
RA	Imose N., Musashino K., Yuuki H., Ohshima A., Sasaki N., Aotsuka S.,
RA	Yoshikawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,
RA	Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,
RA	Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
RA	Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
RA	Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
RA	Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujizumi Y.,
RA	Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,
RA	Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,

Query Match 64.5%; Score 838; DB 2; Length 704;
 Best Local Similarity 61.0%; Pred. No. 2.1e-66;
 Matches 150; Conservative 43; Mismatches 49; Indels 4; Gaps 1;

QY 1 IDGLRNIIWIKPAKSRGRDIVCMRDVREIIEAADAHPILSRDNKVVQKVIETPLLICD 60
 DB 243 MEGDRIWIVKPGAKSRGRGIMCMNRLDMLKLVDCNPMKDKGKVIQKVIETPLLIIFG 302

QY 61 TKFEDIRQWFLVTDWNPPLTWYFVYKESYLRFPSTQRFSLDKLDSAIHLNNVQKVLKNDVGR 120
 DB 303 TKFDLRQWFLVTDWNPPLTWYFVYKESYLRFPSTQRFSLDKLDSVHLNNSIQHLEASCHR 362

QY 121 SPLLPAHNMWTSRFOEYLQRCRGAVGWSVLYPSMKKAIAMKVAODHVPKNSFEL 180
 DB 363 HPMLPPDNNWSSQRFQAHLEQVDAKAWSSVLYPGKAAVTHALQTSQDNVQCRKASFEL 422

QY 181 YGADFVLGRDFRPWLIENSSPTMHPSTVTAQLCAQVOEDTIKAVAV-----DRSCDIGNF 236
 DB 423 YGADFVFGEDFQWLIENASPTMAPSTAVTARLCAGVQADTLRVVIDRDRSCDTGAF 482

QY 237 ELLWRQ 242
 DB 483 ELIYKQ 488

RESULT 10
 Q922T0 MOUSE
 ID Q922T0_MOUSE PRELIMINARY; PRT; 266 AA.
 AC Q922T0;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE 4833441J24Rik protein.
 GN Name=4833441J24Rik;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
 OC Muridea; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=FVB/N; TISSUE=Mammary tumor. C3;
 RX MEDLINE=23288257; PubMed=1477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
 RA Blakesley R.W., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Rodriguez A.C., Trinchum J.W., Green E.D., Dickson M.C.,
 RA Butterfield V.S.N., Krzywinski M.I., Skalska U., Smalusz D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=FVB/N; TISSUE=Mammary tumor. C3;
 RG NIH MGC Project;
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC006830; AA06830.1; -; mRNA.
 DR Ensembl; ENSMUSG0000030276; Mus musculus.
 DR MGI; MGI:2141418; 4833441J24Rik.
 DR GO; GO:0004835; F:tubulin-tyrosine ligase activity; IEA.
 DR GO; GO:0006464; P:protein modification; IEA.

DR InterPro; IPR004344; Tub_tyr_ligase.
 DR Pfam; PF03133; TTL; 1.
 SQ SEQUENCE 266 AA; 30506 MW; F8B8FB52FA8B8E98 CRC64;

Query Match 64.2%; Score 834; DB 2; Length 266;
 Best Local Similarity 60.6%; Pred. No. 1.5e-66;
 Matches 149; Conservative 44; Mismatches 49; Indels 4; Gaps 1;

QY 1 IDGLRNIIWIKPAKSRGRDIVCMRDVREIIEAADAHPILSRDNKVVQKVIETPLLICD 60
 DB 1 MEGDRIWIVKPGAKSRGRGIMCMNRLDMLKLVDCNPMKDKGKVIQKVIETPLLIIFG 60

QY 61 TKFEDIRQWFLVTDWNPPLTWYFVYKESYLRFPSTQRFSLDKLDSAIHLNNVQKVLKNDVGR 120
 DB 61 TKFDLRQWFLVTDWNPPLTWYFVYKESYLRFPSTQRFSLDKLDSVHLNNSIQHLEASCHR 120

QY 121 SPLLPAHNMWTSRFOEYLQRCRGAVGWSVLYPSMKKAIAMKVAODHVPKNSFEL 180
 DB 121 HPMLPPDNNWSSQRFQAHLEQVDAKAWSSVLYPGKAAVTHALQTSQDNVQCRKASFEL 180

QY 181 YGADFVLGRDFRPWLIENSSPTMHPSTVTAQLCAQVOEDTIKAVAV-----DRSCDIGNF 236
 DB 181 YGADFVFGEDFQWLIENASPTMAPSTAVTARLCAGVQADTLRVVIDRDRSCDTGAF 240

QY 237 ELLWRQ 242
 DB 241 ELIYKQ 246

RESULT 11
 Q58CT2 BOVIN
 ID Q58CT2_BOVIN PRELIMINARY; PRT; 261 AA.
 AC Q58CT2;
 DT 10-MAY-2005 (TREMBlrel. 30, Created)
 DT 10-MAY-2005 (TREMBlrel. 30, Last sequence update)
 DT 10-MAY-2005 (TREMBlrel. 30, Last annotation update)
 DE Tubulin tyrosine ligase-like family, member 3.
 GN Name=TTLL3;
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
 OC Pecora; Bovidae; Bovinae; Bos.
 OC NCBI_TaxID=9913;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC TISSUE=Pooled;
 RX MEDLINE=21180013; PubMed=11282978; DOI=10.1101/gr.170101;
 RA Smith T.P.L., Grosse W.M., Freking B.A., Roberts A.J., Stone R.T.,
 RA Casas E., Wray J.E., White J., Cho J., Fahrenkrug S.C., Bennett G.L.,
 RA Heaton M.P., Laegreid W.W., Rohrer G.A., Chitko-McKown C.G.,
 RA Petrea G., Holt I., Karamycheva S., Liang F., Quackenbush J.,
 RA Keele J.W.;
 RT "Sequence evaluation of four pooled-tissue normalized bovine cDNA
 RT libraries and construction of a gene index for cattle."
 RL Genome Res. 11:626-630 (2001).
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RC TISSUE=Pooled;
 RA Hathay G.P., Sonstegard T.S., Van Tassell C.P., Clawson M.L.,
 RA Heaton M.P., Keele J.W., Snelling W.M., Weidmann R.T., Smith T.P.L.;
 RT "Sequencing and analysis of Bos taurus full-length insert cDNA
 RT clones."
 RL Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BT021865; AAX46712.1; -; mRNA.
 KW Ligase.
 SQ SEQUENCE 261 AA; 29911 MW; 5F069784CA162017 CRC64;

Query Match 61.5%; Score 799; DB 2; Length 261;
 Best Local Similarity 61.6%; Pred. No. 2e-63;
 Matches 143; Conservative 37; Mismatches 52; Indels 0; Gaps 0;

QY 1 IDGLRNIIWIKPAKSRGRDIVCMRDVREIIEAADAHPILSRDNKVVQKVIETPLLICD 60
 DB 1 MEGDRIWIVKPGAKSRGRGIMCMNRLDMLKLVDCNPMKDKGKVIQKVIETPLLIIFG 60

```

Db 3 MEGDRNIWIVKPGAKSRGRCIMCWDHLEBMLKLVDNCNPMWMDGKWVHNYIERPLLIIG 62
Qy 61 TKFDIRQWFLVTDWNPPLTTFWFKESYLRSTQRFSLDKLDSAIHLCCNNAVOKYKNDVGR 120
Db 63 TKFDLRQWFLVTDWNPPLTTFWFKESYLRSTQRFSLDKLDSAIHLCCNNAVOKYKNDVGR 122
Qy 121 SPLPAHNNWTSRFOEYLQROGRGAVGWSVLYPSMKKAIAHAMKVAQDHVPRKNSFEL 180
Db 123 HPLLPPDNWSSQFQAHQETGAPNAWSTVIVPGMKAAVIHALQTSQDVTQCRKASFEL 182
Qy 181 YGADFVLGRDFRPLWLEINSPTMHPSTPVTQAOLCAQVQEDTIKVAVDKSCD 232
Db 183 YGADFVFGEDFQWLEINASPTMAPSTAVTARLCAGVQADTLRVVIDWRLD 234

RESULT 12
Q4RY08_TETNG
ID Q4RY08_TETNG PRELIMINARY; PRT; 534 AA.
AC Q4RY08;
DT 13-SEP-2005 (T-EMBLrel. 31, Created)
DT 13-SEP-2005 (T-EMBLrel. 31, Last sequence update)
DE 13-SEP-2005 (T-EMBLrel. 31, Last annotation update)
DE Chromosome 11 SCAF14979, whole genome shotgun sequence.
GN ORFNames=GSTENG0027209001;
OS Tetraodon nigroviridis (green puffer).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Perciformes; Tetraodontiformes;
OC Tetraodontidae; Tetraodontidae; Tetraodon.
OX NCBI_TaxID=99883;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Jaillon O., Aury J.M., Brunet F., Petit J.L., Stange-Thomann N.,
RA Jaillon O., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
RA Mauceli E., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
RA Anthouard V., Jubin C., Castellani V., Katinka M., Vacherie B.,
RA Biemont C., Skalli Z., Cottolero L., Poullain J., De Berardinis V.,
RA Craud C., Duprat S., Brottier P., Coutanceau J.P., Gouzy J.,
RA Parra G., Lardier G., Chapple C., McKernan K.J., McEwan P., Bosak S.,
RA Kellis M., Wolff J.N., Guigo R., Zody M.C., Mesirov J.,
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
RA Wincker P., Lander E.S., Weissbach J., Roest Crollius H.;
RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
RL the early vertebrate proto-karyotype.";
RN Nature 431:946-957(2004).
RP [2]
RG Genoscope; Whitehead Institute Centre for Genome Research;
RG Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; CAE01014979; CAG06724.1; -; Genomic DNA.
DR EMBL; 534 AA; 60130 MW; 62B5B5B696B62B CRC64;
SQ SEQUENCE 534 AA; 60130 MW; 62B5B5B696B62B CRC64;

Query Match 60.5%; Score 786; DB 2; Length 534;
Best Local Similarity 59.2%; Pred. No. 7e-62;
Matches 145; Conservative 36; Mismatches 60; Indels 4; Gaps 1;

Qy 2 DGLRNIWIKPAKSRGRDIVCMRDVBEILEAAADHPLSRDNKVVQKYIETPLLCDT 61
Db 262 DGLNNIWIKPGAMSRGIVCKDLDEILAVDNDKSLNKKKKVQKYLERPLLVHGT 321
Qy 62 KFDTRQWFLVTDWNPPLTTFWFKESYLRSTQRFSLDKLDSAIHLCCNNAVOKYKNDVGRS 121
Db 322 KFDLRQWFLVTDWNPPLTTFWFKESYLRSTQRFSLDKLDSAIHLCCNNAVOKYKNDVGRS 121
Qy 122 PLLPAHNNWTSRFOEYLQROGRGAVGWSVLYPSMKKAIAHAMKVAQDHVPRKNSFELY 181
Db 382 PAVPGDNWSSQFQAHQETGAPNAWSTVIVPGMKAAVIHALQTSQDVTQCRKASFELY 441

```

```

Qy 182 GADFVLGRDFRPLWLEINSPTMHPSTPVTQAOLCAQVQEDTIKVAVDKSCD ----IGNFE 237
Db 442 GADFVLGRDFRPLWLEINSPTMHPSTPVTQAOLCAQVQEDTIKVAVDKSCD ----IGNFE 237
Qy 238 LLWRQ 242
Db 502 LICRQ 506

RESULT 13
Q7Q156 ANOQA
ID Q7Q156 ANOQA PRELIMINARY; PRT; 572 AA.
AC Q7Q156;
DT 01-MAR-2004 (T-EMBLrel. 26, Created)
DT 01-MAR-2004 (T-EMBLrel. 26, Last sequence update)
DE 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)
DE ENSANGP0000022337 (Fragment).
GN ORFNames=ENSANG0000019848;
OS Anopheles gambiae str. PEST.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Culicidae;
OC Anophelinae; Anopheles.
OX NCBI_TaxID=180454;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA STRAIN=PEST;
RG The Anopheles gambiae Sequence Committee;
RT "Anopheles gambiae re-annotation.";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RA STRAIN=PEST;
RG The Anopheles gambiae Sequence Committee;
RL Submitted (APR-2004) to the EMBL/GenBank/DBJ databases.
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAB01008980; EAA13905.2; -; Genomic DNA.
DR GO; GO:0004835; Fubulin-tyrosine ligase activity; IEA.
DR GO; GO:0006464; P:protein modification; IEA.
DR InterPro; IPR004344; Tub_tyr_lygase.
DR Pfam; PF03133; TTL; 1.
FT NON_TER 1 1
FT NON_TER 572 572
SQ SEQUENCE 572 AA; 67109 MW; 8938D9EDD5935071 CRC64;

Query Match 49.6%; Score 644.5; DB 2; Length 572;
Best Local Similarity 50.4%; Pred. No. 4e-49;
Matches 124; Conservative 45; Mismatches 70; Indels 7; Gaps 3;

Qy 1 IDGLRNIWIKPAKSRGRDIVCMRDVBEILEAAADHPLSRDNKVVQKYIETPLLCD 60
Db 305 LDGLYNIWIVKPGNKGKRGHLMNNIKQI--TAMVNPPIVSKTRVYQKYIERPLIHN 362
Qy 61 TKFDIRQWFLVTDWNPPLTTFWFKESYLRSTQRFSLDKLDSAIHLCCNNAVOKYKNDVGR 120
Db 363 TKFDIRQWFLVTDWNPPLTTFWFKESYLRSTQRFSLDKLDSAIHLCCNNAVOKYKNDVGR 120
Qy 121 SPLPAHNNWTSRFOEYLQROGRGAVGWSVLYPSMKKAIAHAMKVAQDHVPRKNSFEL 180
Db 422 DERLPENWDCHTFQAYLRQIDKYENWSEIRIVPGMKAIIGSILLACQDNNRRPNTFEL 481
Qy 181 YGADFVLGRDFRPLWLEINSPTMHPSTPVTQAOLCAQVQEDTIKVAVDKSCD ----IGNF 236
Db 482 YGADFMTTFDYPWLEINSPTLAPSTSVTARLCPCQVEDTIRVIDRRTDSNAPTGSF 541
Qy 237 ELLWRQ 242
Db 542 ELIYKQ 547

RESULT 14
Q9VM91_DROME

```

ID AC Q9VM91_DROME PRELIMINARY; PRT; 992 AA.
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE CG11323-PA.
 GN Name=CG11323; ORFNames=CG11323;
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1] _TaxID=7227;
 RN NUCLEOTIDE SEQUENCE.
 RP MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazef R.G., Champagne M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burris K.C., Busam D.A., Butler H., Cadiou E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko F., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirska R.A., Tector C., Turner C., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195 (2000).
 RN [2]
 RN NUCLEOTIDE SEQUENCE.
 RP MEDLINE=22426065; PubMed=12537568;
 RA Celniker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,
 RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
 RA George R.A., Hoskins R.A., Lavery T., Muzny D.M., Nelson C.R.,
 RA Pacle J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
 RA Svirska R.A., Taber P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
 RA Weinstein G., Scher S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
 RT "Finishing a whole-genome shotgun: release 3 of the *Drosophila*
 RT melanogaster euchromatic genome sequence.";
 RL Genome Biol. 3:RESEARCH0079-RESEARCH0079 (2002).
 RN [3]
 RN NUCLEOTIDE SEQUENCE.
 RP MEDLINE=22426070; PubMed=12537573;
 RA Kaminker J.S., Bergman C.M., Krommiller B., Carlson J.W., Svirska R.,
 RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
 RA Ashburner M., Celniker S.E.;
 RT "The transposable elements of the *Drosophila melanogaster* euchromatin:
 RT a genomics perspective."

RL Genome Biol. 3:RESEARCH0084.1-RESEARCH0084.20 (2002).
 RN [4]
 RN NUCLEOTIDE SEQUENCE.
 RX MEDLINE=22426069; PubMed=12537572;
 RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
 RA Hradecky P., Huang Y., Kaminker J.S., Milburn G.H., Prochuk S.E.,
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
 RA Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
 RA Lewis S.E.;
 RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a
 RT systematic review.";
 RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22 (2002).
 RN [5]
 RN NUCLEOTIDE SEQUENCE.
 RP Berkeley Drosophila Genome Project;
 RA Celniker S., Carlson J., Wan K., Pfeiffer B., Frise E., George R.,
 RA Hoskins R., Stapleton M., Pacle J., Park S., Svirska R., Smith E.,
 RA Yu C., Rubin G.;
 RT "Drosophila melanogaster release 4 sequence.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBSJ databases.
 RN [6]
 RN NUCLEOTIDE SEQUENCE.
 RP FlyBase;
 RL Submitted (MAR-2005) to the EMBL/GenBank/DBSJ databases.
 DR EMBL; AE003614; AAF52432.1; -: Genomic DNA.
 DR Ensembl; CG11323; Drosophila melanogaster.
 DR FlyBase; FBgn0031854; CG11323.
 DR GO; GO:0004835; F: tubulin-tyrosine ligase activity; IEA.
 DR GO; GO:0006464; P: protein modification; IEA.
 DR InterPro; IPR004344; Tub_tyr_lygase.
 DR Pfam; PF03133; TTL; 1.
 SQ SEQUENCE 992 AA; 112255 MW; 0FC34674C15BDAAD CRC64;
 Query Match 48.4%; Score 628.5; DB 2; Length 992;
 Best Local Similarity 48.4%; Pred. No. 2.1e-47;
 Matches 119; Conservative 50; Mismatches 70; Indels 7; Gaps 3;
 QY 1 IDGNWIIIPAKSRGRDVCMDRVEELLEAAADHPLSRDKNVVKYIETPLICD 60
 DB 411 LDGQNMWIVDPANKRGRIILMDNLKKI--LGVVNLISAKSRVYVQKYIERPLIFQ 468
 QY 61 TKFDIOWFLVTDNPLTIWFYKESYLFSTQRFSLDKLDSAIHLCNNAVQKYLNDVGR 120
 DB 469 TKFDIOWFLITNTQPLVWYRESYLFSSQVSLSNHSHVLTNTVAIQKTYNG-KR 527
 QY 121 SPLLPANHWMTSTRFQVYLRQGRGAVMGSVIYPSMKKAIAHAMKVAQDHVEPRKNSPEL 180
 DB 528 DKRLPSENMDVCSFYALRQIGKYNMWERIFPGMKRAIVGCMLASQENMDRRPTEP 587
 QY 181 YGADVFLGRDPRPWLIEINSSPTWHPSTPTVAQLCAQVQEDTIKAVDRSCD----IGNF 236
 DB 588 FGADFMICENFPWLIEINSSPDLGATTSVTARMCPQCLEDVVKVVIDRRTDPKALGNF 647
 QY 237 ELLWRQ 242
 DB 648 ELAYRQ 653
 RESULT 15
 ID O7PMD3 ANOGA PRELIMINARY; PRT; 501 AA.
 AC O7PMD3;
 DT 01-MAR-2004 (TrEMBLrel. 26, Created)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
 DE ENSANGP0000022334 (Fragment)
 GN ORFNames=ENSANGG00000019845;
 OS Anopheles gambiae str. PEST.
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Nematocera; Culicidae; Culicidae;
 OC Anophelinae; Anopheles.

```
OX NCBI_TaxID=180454;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=PEST;
RG The Anopheles gambiae Sequence Committee;
RT "Anopheles gambiae re-annotation.";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=PEST;
RG The Anopheles gambiae Sequence Committee;
RL Submitted (APR-2004) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
preliminary data.
CC EMBL; AAAB01008980; EAA13971.3; -; Genomic DNA.
DR GO; GO:0004835; F:tubulin-tyrosine ligase activity; IEA.
DR GO; GO:0006464; P:protein modification; IEA.
DR InterPro; IPR011761; ATP_GRASP.
DR InterPro; IPR004344; Tub_cyr_lygase.
DR Pfam; PF03133; TTL; 1.
DR PROSITE; PSS0975; ATP_GRASP; 1.
FT NON_TER 1
FT NON_TER 501
SQ SEQUENCE 501 AA; 58862 MW; 806539A2FD8E76DF CRC64;

Query Match 40.1%; Score 521; DB 2; Length 501;
Best Local Similarity 45.9%; Pred. No. 4.4e-38;
Matches 112; Conservative 41; Mismatches 81; Indels 10; Gaps 4;

QY 1 IDGIRNIWIKPAKSRGRDIVCMDRVVEEILELAADHPLSRDNKVVOKYIETPLLICD 60
DB :|||:||||:| :|| :| :||| :| :||| :||| :||| :|||
262 IDGIRNWIILKPGNCRGGLGIMLFNDNRKLEHVD-----NPDVKYVAQKYIERPLLIHC 317
QY 61 TKFDIROWFLVT-DWNPFLTWFYKESYLRFSTQRFSLDKLDSAIHLCNNAVQK-YLKNDV 118
DB :|||:||||:| :|| :| :||| :||| :||| :||| :||| :|||
318 TKFDIROYFLITITNNVLKVMYRNCYLRFSSRQFNLDLDFSEIHLTNYSIQKNYAKEVR 377
QY 119 GRSPLLPAHNMWTSTPQEYLQGRGAVWGVSIVYPSMKKAIAMKVAQDHVEPRKNSF 178
DB :|||:||||:| :|| :| :||| :||| :||| :||| :||| :|||
378 EGADALPASNNWSLKRFOEHLQSLDKGFYWERKIYPMKKNILAIVCASLDGKMKERNMF 437
QY 179 ELYGADFVLGRDRPMLIEINSSPTMHPSTPTAQLCAQVQEDTIKVADRESCD----IG 234
DB :|||:||||:| :|| :| :||| :||| :||| :||| :||| :|||
438 ELYGADFMVTNFTMLIEINTSPDLSSSTDVTSVICPAVLEDLVKVVIDNTKDKRAGTG 497
QY 235 NFEL 238
DB :|||
498 QFEL 501
```

Search completed: April 4, 2006, 12:45:57
Job time : 36.6599 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2006, 12:45:31 ; Search time 8.99628 Seconds
(without alignments)
2588.231 Million cell updates/sec

Title: US-10-635-977-2_COPY_133_374

Perfect score: 1299
Sequence: 1 IDGLRNIIWIKPAKSRGRD.....IKVAVDRSCDIGNFELLWRQ 242

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: Pir1.*
- 2: Pir2.*
- 3: Pir3.*
- 4: Pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	850	65.4	352	2 T12515	hypothetical prote
2	247.5	19.1	379	2 A45443	tubulin-tyrosine 1
3	243	18.7	640	2 E88575	protein ZK1128.6
4	243	18.7	680	2 T27699	hypothetical prote
5	209.5	16.1	1198	2 T20262	hypothetical prote
6	209.5	16.1	1203	2 C89217	protein C55A6.2
7	156	12.0	662	2 T20343	hypothetical prote
8	155.5	12.0	883	2 A96805	hypothetical prote
9	155	11.9	403	2 T37571	tubulin-tyrosine 1
10	144	11.1	753	2 S48261	hypothetical prote
11	87	6.7	439	2 T01270	hypothetical prote
12	85	6.5	1872	2 T30888	vitellogenin - Ath
13	84.5	6.5	522	2 T45824	hypothetical prote
14	83.5	6.4	374	2 E85920	conserved hypothet
15	83.5	6.4	930	2 A25923	progesterone recep
16	82.5	6.4	349	2 F91218	probable transport
17	82.5	6.4	349	2 H86064	probable transport
18	82.5	6.4	349	2 D65182	probable transport
19	82	6.3	306	2 A25698	probable protein k
20	82	6.3	1124	2 JH0588	calmodulin-binding
21	81.5	6.3	717	2 AC1419	DNA topoisomerase
22	81	6.2	270	2 B84813	probable RING zinc
23	81	6.2	524	2 T43050	cyclin E - Caenorh
24	81	6.2	570	2 T30156	hypothetical prote
25	80	6.2	1577	2 T30858	glucosyltransferas
26	80	6.2	2176	2 T39188	probable U5 snRNP
27	79.5	6.1	380	2 H81152	anticodon nuclease
28	79.5	6.1	433	2 T04594	aldehyde dehydroge
29	79.5	6.1	1707	2 S77908	hypothetical prote

30	79.5	6.1	2142	2 D86303	P17F16.1 protein -
31	79	6.1	733	2 A64650	flagellar biosynth
32	79	6.1	733	2 E71937	flagellar biosynth
33	79	6.1	742	2 F84643	hypothetical prote
34	79	6.1	1528	2 S13743	DNA strand transe
35	78.5	6.0	286	2 G69537	ribosomal protein
36	78.5	6.0	478	2 T03750	violaxanthin de-ep
37	78.5	6.0	515	2 C71158	probable thermosta
38	78.5	6.0	857	2 T04208	probable anthranil
39	78.5	6.0	1139	2 T33368	hypothetical prote
40	78	6.0	1012	2 S68259	DNA polymerase gam
41	77.5	6.0	627	2 AE2714	hypothetical prote
42	77.5	6.0	637	2 B97496	hypothetical prote
43	77.5	6.0	751	1 TVVPTH	large T antigen -
44	77.5	6.0	950	2 A82986	adenylate cyclase
45	77	5.9	322	2 AF2725	transcription regu

ALIGNMENTS

RESULT 1

T12515
hypothetical protein DKFZp434B103.1 - human
C:Species: Homo sapiens (man)
C>Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 09-Jul-2004
C:Accession: T12515
R:Wambutt, R.; Heubner, D.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, June 1999
A:Reference number: Z17524
A:Accession: T12515
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-352 <WAM>
A:Cross-references: UNIPROT:O9YAR7; UNIPARC:UPI00001377A6; EMBL:AL096725
A:Experimental source: adult testis; clone DKFZp434B103
C:Genetics:
A>Note: DKFZp434B103.1

Query Match	65.4%	Score 850;	DB 2;	Length 352;
Best Local Similarity	61.8%	Pred. No. 7.1e-71;		
Matches 152;	Conservative 41;	Mismatches 49;	Indels 4;	Gaps 1;
Qy	1	IDGLRNIIWIKPAKSRGRDIVCMRDVVEIELEAAADHPLSRDNKVVQKVIETPLLICD	60	
Db	61	MEGDRIWIKPGAKSRGIMCDHLEMLKLVNGNPVWKGKGVQKVIETPLIFG	120	
Qy	61	TKFDIRQWFLVDWNPITWFKESYLRFSTQRFSLDKLDSAIHLCNNAVQKYLKNDVGR	120	
Db	121	TKFDLRQWFLVDWNPITWFKESYLRFSTQRFSLDKLDSAIHLCNNAVQKYLENSCHR	180	
Qy	121	SPLLPAHNMWTSRFOELQQRGAVGWSVIYPSMKKAIAMKVAODHVEPRKNSFEL	180	
Db	181	HPLLPDNNWSQRFQALHQMGNAPNAWSTIIVPGMDKAVIHALQTSQDTVQCRKASFEL	240	
Qy	181	YGADFVLGRDPRPWLIIENSPTMHPSTPTVAQLCAQVQEDTIKIVAV----DRSCDIGNF	236	
Db	241	YGADFVGEDPQPLIENSPTMAPSTAVTARLCAGVQADTLRVVIDRMLDRNCOTGAF	300	

RESULT 2

A45443
tubulin-tyrosine ligase (EC 6.3.2.25) [validated] - pig
C:Species: Sus scrofa domestica (domestic pig)
C>Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
C:Accession: A45443
R:Rersfeld, K.; Wenland, J.; Plessmann, U.; Dodemont, H.; Gerke, V.; Weber, K.
J. Cell Biol. 120, 725-732, 1993
A:Title: Characterization of the tubulin-tyrosine ligase.

A;Reference number: A45443; MUID:93147125; PMID:8093886
A;Accession: A45443
A;Status: preliminary
A;Molecule type: mRNA; protein
A;Residues: 1-379 <ERS>
A;Cross-references: UNIPROT:P38160; UNIPARC:UPI00001377A9; GB:X68453; GB:S54050; NID:521
A;Experimental source: brain
A;Note: sequence extracted from NCBI backbone (NCBIN:123858, NCBIP:123859)
C;Keywords: ligase

Query Match 19.1%; Score 247.5; DB 2; Length 379;
Best Local Similarity 29.4%; Pred. No. 4.4e-15;
Matches 69; Conservative 40; Mismatches 93; Indels 33; Gaps 9;

QY 2 DGLRNIWIKPAKSRGRDIVCMRDVEEILELAADHPILSRDNK---WVQKYIETPLLI 58
DB 141 DGEVNWIAKSSAGAKGEGILISSEATELLDFI-----DNQGVHVQKYLRLPLLL 192
QY 59 --CDTKFIQWFLVTDWNPITTFYKESYLRFPSTORFSLDKL-DSAIHLCNNAVOKYLYK 115
DB 193 EPGHRKFDIRSVLVD--HQYNIYLYREGVLRFTASEPYHTDNFQDKTCHLTNHCIOKEYS 250
QY 116 NDVGRSPLPAHNMWTSRFOEYLQGRGAVGWSVIYPSMKKAIAHAMKVAQDHVEPR- 174
DB 251 KNYGK---YEEGNEMFEEFNQVLT-----SALNITLESILLOIKHIIIRSCLLSVEPAI 302
QY 175 -----KNSPELYGADFVLGRDPRPWLIEINSSPTMHPSTPVTQAQLCAQVQEDTI 223
DB 303 STRHLPYQSOLFQGFDFWDEDLKWLIEVNGAPAC--AQKLYAELCQGIVDIAI 355

RESULT 3
E88575
protein ZK1128.6 [imported] - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
R;anonymous, The C. elegans Sequencing Consortium.
C;Accession: E88575
A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
A;Reference number: A75000; MUID:99069613; PMID:9851916
A;Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C_elegans/
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A;Accession: E88575
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-640 <STO>
A;Cross-references: UNIPROT:Q09647; UNIPARC:UPI000017A077; GB:chr_III; PIDN:CAA87425.1;
C;Genetics:
A;Gene: ZK1128.6
A;Map position: 3

Query Match 18.7%; Score 243; DB 2; Length 640;
Best Local Similarity 27.8%; Pred. No. 2.3e-14;
Matches 67; Conservative 47; Mismatches 89; Indels 38; Gaps 9;

QY 2 DGLRNIWIKPAKSRGRDIVCMRDVEEILELAADHPILSRDNKWWVQKYIETPLLICDT 61
DB 285 DASHRV-IVKPPASARGTGISVTRKPKDFPTATL-----VAQHVIERPLTINRA 333
QY 62 KFDIROWFLVTDWNPITTFYKESYLRFPSTORFSLDK---LDSAIHLCNNAVOKYLYKND- 117
DB 334 KFDRLYAYVPTFEPLRVYIDQGLVRFAVPYSHSVSTISNKYMLTNYSinkLAEADG 393
QY 118 VGRSPL--LPAHNMWTSRFP-----QEYLQGRGAVGWSVIYPSMKKAIAHAMKVAQD 169
DB 394 VANKPVPKWTLLHLW--EHPDEMGVDREKIQRE-----IEEVIKAFISTEKPIRE 442
QY 170 H----VPRKNSPELYGADFVLGRDPRPWLIEINSSPTMHPSTPVTQAQLCAQVQEDTIKV 225
DB 443 HNSRFLQEFCYELFGIIDLDEYKFWLLEVNISPSLHSGTFLDVSVKAPLAKDVLNL 502
QY 226 A 226

Db 503 A 503

RESULT 4
T27699
hypothetical protein ZK1128.6 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T27699
R;Berks, M.
submitted to the EMBL Data Library, January 1995
A;Reference number: Z20407
A;Accession: T27699
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-680 <WIL>
A;Cross-references: UNIPROT:Q09647; UNIPARC:UPI000017A076; EMBL:Z47357; PIDN:CAA87425.2
A;Experimental source: clone ZK1128
C;Genetics:
A;Gene: CESP:ZK1128.6
A;Map position: 3
A;Introns: 114/1; 208/3; 242/3; 334/3; 362/2; 405/2; 424/2; 550/2; 633/2

Query Match 18.7%; Score 243; DB 2; Length 680;
Best Local Similarity 27.8%; Pred. No. 2.5e-14;
Matches 67; Conservative 47; Mismatches 89; Indels 38; Gaps 9;

QY 2 DGLRNIWIKPAKSRGRDIVCMRDVEEILELAADHPILSRDNKWWVQKYIETPLLICDT 61
DB 325 DASHRV-IVKPPASARGTGISVTRKPKDFPTATL-----VAQHVIERPLTINRA 373
QY 62 KFDIROWFLVTDWNPITTFYKESYLRFPSTORFSLDK---LDSAIHLCNNAVOKYLYKND- 117
DB 374 KFDRLYAYVPTFEPLRVYIDQGLVRFAVPYSHSVSTISNKYMLTNYSinkLAEADG 433
QY 118 VGRSPL--LPAHNMWTSRFP-----QEYLQGRGAVGWSVIYPSMKKAIAHAMKVAQD 169
DB 434 VANKPVPKWTLLHLW--EHPDEMGVDREKIQRE-----IEEVIKAFISTEKPIRE 482
QY 170 H----VPRKNSPELYGADFVLGRDPRPWLIEINSSPTMHPSTPVTQAQLCAQVQEDTIKV 225
DB 483 HNSRFLQEFCYELFGIIDLDEYKFWLLEVNISPSLHSGTFLDVSVKAPLAKDVLNL 542
QY 226 A 226
DB 543 A 543

RESULT 5
T20262
hypothetical protein C55A6.2 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T20262
R;Kershaw, J.
submitted to the EMBL Data Library, October 1996
A;Reference number: Z19243
A;Accession: T20262
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-1198 <WIL>
A;Cross-references: UNIPROT:O17720; UNIPARC:UPI000017A066; EMBL:Z81051; PIDN:CAB02862.2
A;Experimental source: clone C55A6
C;Genetics:
A;Gene: CESP:C55A6.2
A;Map position: 5
A;Introns: 32/3; 68/1; 107/3; 201/3; 273/1; 364/3; 407/3; 460/3; 476/2; 534/2; 677/3; 7

Query Match 16.1%; Score 209.5; DB 2; Length 1198;
Best Local Similarity 25.5%; Pred. No. 6.4e-11;
Matches 60; Conservative 57; Mismatches 87; Indels 31; Gaps 9;

Qy	8	WIIKPAKSRGRDIVCMQDRVZEIILEAAADHPLSRDNKWVQKYIETPELLICDTKFDIRQ	67
Db	715	FIVEPTNSRQKGIFFANSMADI-----PAEGPL-----LVSRYLKDPYLVNNHFXDLRI	764
Qy	68	WFLVTDWNPFLTIWFKESYLRFSTORF--SLDKLDS--AIHLCNNAVOK-----YLKNDVG	119
Db	765	YVATVSFYPLVAYVYSEGLARLAGRPYDTSSADSNSEYVHLTNTYSINKNSTSFVYRNESM	824
Qy	120	RSPLLPAAHNMWTSTRFOEYLQRQGRGAVMGSVIYPSMKKAIAHAKVAQDHY--BPRKN-	176
Db	825	SSEDL--GHKWTIGALLRYVNEGKOA--XLLMLRIEDLVKSLLSIONSVAATSRINL	879
Qy	177	-----SFELYGADFVLGRDFRPWLIEINSSPTMHSTPVTVAQLCAQVOEDTIKVA	226
Db	880	RFACTNFELFGDVLVDQALKEPVLLEVNLSPLACDAPLDSLLKTRLLADILNLA	934

RESULT 6

C89217
protein C55A6.2 [imported] - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
C;Accession: C89217
R;anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
A;Reference number: A75000; PMID:98069613; PMID:9851916
A;Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/projects/c.ele
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A;Accession: C89217
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1203 <STO>
A;Cross-references: UNIPROT:O17720; UNIPARC:UPI000017A067; GB:chr_V; PIDN:CAB02862.1; PI
C;Genetics:
A;Gene: C55A6..2
A;Map position: 5

Query Match	16.1%	Score	209.5	DB 2	Length	1203			
Best Local Similarity	25.5%	Pred. No.	6.5e-11						
Matches	60	Conservative	57	Mismatches	87	Indels	31	Gaps	9

QY	8	WIIPAAKSRGRD	IVCMDRV	EEILELAAAD	PLSRDNK	WVYQKY	ETETLLIC	DTKFD	IURQ	67																																						
					:	:	:	:																																								
DB	720	FIVKPTNSRQ	KGIFFANS	MADI	----	PAEGPL	-----	LVS	RYLKOP	YLVNHNKFD	LR	769																																				
				:	:	:	:	:	:		:																																					
QY	68	WFLVTDN	PLTIN	WPKSY	LR	PFSTOR	----	SLDKLDS	----	AHL	CNN	AVOK	----	YLNK	DVG	119																																
				:	:	:	:		:	:	:	:	:	:	:	:																																
DB	770	YVAWTS	FYPL	VAYV	YSEGL	ARLAS	RPDY	TSASS	ADSNE	YVHL	TNY	SINK	NSTS	FV	RNE	SM	829																															
				:	:	:	:	:	:	:	:	:	:	:	:	:																																
QY	120	RSPLL	PAHN	MW	TS	TRFQ	EYLR	QGR	GA	WGS	VI	PS	MKA	A	HA	M	KA	Q	D	H	V	----	E	P	R	K	N	----	176																			
				:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:																		
DB	830	SSEDL	----	CHK	WT	GALL	RY	VEN	E	K	D	----	KL	M	R	I	E	D	L	I	V	K	S	L	L	T	S	T	Q	N	S	V	A	T	S	R	N	L	----	884								
				:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:									
QY	177	-----	SPE	L	Y	A	D	F	V	L	G	R	F	P	M	L	I	N	S	S	T	M	H	P	S	T	P	T	V	T	A	Q	L	C	A	Q	V	O	E	D	I	T	K	V	A	----	226	
				:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	
DB	885	RFAC	TN	PE	L	G	F	O	V	I	D	Q	A	L	K	P	W	L	E	N	L	S	P	S	L	A	C	D	A	P	L	D	S	L	L	K	T	R	I	A	D	I	L	N	L	A	----	939

RESULT 7

T20343
hypothetical protein D2013.9 - *Caenorhabditis elegans*
C/Species: *Caenorhabditis elegans*
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C/Accession: T20343; T22085
R/Mortimore, B.
submitted to the EMBL Data Library, January 1995
A/Reference number: Z19259
A/Accession: T20343
A/Status: preliminary; translated from GB/EMBL/DDBJ
A/Molecule type: DNA
A/Residues: 1-662 <WIL>
A/Cross-references: UNIPROT: Q099512; UNIPARC: UPT000013B847; EMBL: Z47808; PIDN: CAA87778.1;

A;Experimental source: clone D2013
R;Matthews, P.
submitted to the EMBL Data Library, January 1995
A;Reference number: Z19510
A;Accession: T22085
A;Status: preliminary;
A;Molecule type: DNA
A;Residues: 1-662 <W12>
A;Cross-references: UNIPARC:UPI000013B847; EMBL:Z47122
A;Experimental source: clone F42A8
C;Genetics:
A;Gene: CESP:D2013.9
A;Map position: 2
A;Introns: 78/2; 114/3; 330/3; 471/1; 521/2; 587/2

Query Match	12.0%	Score 156;	DB 2;	Length 662;
Best Local Similarity	25.8%;	Pred. No. 2.7e-06;		
Matches	62;	Conservative 36;	Mismatches 100;	Indels 42; Gaps 11;
QY	1	IDGLRIWIIPKPAAKGRGDRIVCMQDVRVEETLEAAADHPLSRDNKWVVKVIETPLLI--	58	
Db	431	LNGCHNVWIKPWNLAGMDMTVEDLNQIRM-----FTGPKVICVYIPRPLIFPR	483	
QY	59	----CDTKEDIRQWFLVTOWNPLTIWFYKESYLRFSTQRFSLKL-DSA-IHLGNNAVOKY	113	
Db	484	PDNGNKKVPDLRYIVFLNGIAPTVAYVNFRFWTRFAINEFSLSNFDVETHF---TVFN	540	
QY	114	LKNDDVGRSPLPAHNMMWTSTRFOEYLQRCRGAVMGSV---IYPSMKAL-AHAMKVAQD	169	
Db	541	L--DKEKILQMCKEN-----FIETIEKAYPRIQNSEVQKDINLTIKATEAANKKEEAPR	592	
QY	170	HVEPRKNSPELYGADFVLGRD-----FRPWLIETINSSTMPHPSTVTAQLCAQVQE--DTI	223	
Db	593	GVAPNVQSRAHYGVDMLOHGONDVITKSTLLEINFMPD-----TTRACQYPDFADTV	645	

RESULT 8

A96805
hypothetical protein TSM16.14 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C:Accession: A96805
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.
ansin, N.F.; Hughes, B.; Huizarr, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, S.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: A96805
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-883 <STO>
A:Cross-references: UNIPROT:Q9CAQ1; UNIPARC:UPI000000A09A3; GB:AE005173; NID:g6382502; i
C:Genetics:
A:Gene: TSM16.14
A:Map position: 1

	Query Match	12.0%;	Score 155.5;	DB 2;	Length 883;	
	Best Local Similarity	25.3%;	Pred. No. 4.3e-06;			
	Matches	55;	Conservative 36;	Mismatches 93;	Indels 33;	Gaps 7;
QY	2	DGURNIWKPAKGRDIVCMCDRVBEELAAADHPLSRONKWVWVKYKIETPLICDT	61			
	:	:	:	:	:	:
	:	:	:	:	:	:
Db	641	DQLNNLWILKPWNARTIDTSITDNLSAIR-----MMETGPCKIQKYIEHPALFKGN	693			
	:	:	:	:	:	:
	:	:	:	:	:	:
QY	62	KFDIQGFVLVTOWNPLTIWFYKESYLRFPTQSFSLDKLSAHLCHNNAVKYIKNDVGRS	121			
	:	:	:	:	:	:
	:	:	:	:	:	:
Db	694	KFDLRVVILVRSIDPLELYLIIFWRLSNNPYSLEKHSFFFEYETHFTVWNY-----GRK	748			
	:	:	:	:	:	:
	:	:	:	:	:	:

QY 122 PLLPAHNMWTSRFOEYLQRG-----RGAVGWSV-----IYPSMK-----KAIAHAMK 165
Db 749 -----LNHRKPAEVRFBFEQBNDFYAFHFVNNTIQLSIVKWMIDHEKVKQVIRAVFEAAA 804
QY 166 VAQDHVEPRKNSFELYGADPVLGRDRFRPMLIEINSSP 202
Db 805 LAHPMQSPK-SRAMYGVDMVLDSSPEPKILEVTCYCP 840

RESULT 9
T37571
tubulin-tyrosine ligase related protein - fission yeast (Schizosaccharomyces pombe)
C/Species: Schizosaccharomyces pombe
C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C/Accession: T37571
R;/Badcock, K.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
submitted to the EMBL Data Library, April 1996
A/Reference number: 221726
A/Accession: T37571
A/Status: preliminary; translated from GB/EMBL/DDBJ
A/Molecule type: DNA
A/Residues: 1-403 <BAD>
A/Cross-references: UNIPROT:Q10438; UNIPARC:UPI000013A94B; EMBL:X70721; PIDN:CAA94694.1;
A/Experimental source: strain 972h-; cosmid c12B10
C/Genetics:
A/Gene: SPDB:SPAC12B10.04
A/Map position: 1
A/Introns: 320/3; 348/3

Query Match 11.9%; Score 155; DB 2; Length 403;
Best Local Similarity 21.3%; Pred. No. 1.7e-06;
Matches 54; Conservative 53; Mismatches 90; Indels 56; Gaps 10;

QY 8 WIIPKPAKSRGRDIVCMRVEIELEL-----AAADHPILSRDNK----- 45
Db 144 YILKPSMCDRAQGRILFSFTIEELQAIFDSFDDSESESEAGLEKEGDIITVAFNNKIVISQ 203
QY 46 ---WVQKYIETPLLICDTKFDIRQWFLVTDNMPLTIWFKYESYLRFSTQRFSLDKLDSA 102
Db 204 IRNFLVQKYISKPLLLDHRKHFIRAYVLAT--GALSVYLENEMLCILARDKYKKPTDPD 261
QY 103 I---HLCNNAVQKYLKNDVGRSPLLPANHNMWTSRFOEYLQRGAGVGSVLYPSMKKA 159
Db 262 LLFSLHSNTCLQ---GDNVQSST---RDFWNTS-----IENK-----DDIPKSLINI 303
QY 160 IAHAMKVAQD---HVEPRKNSFELYGADPVLGRDRFRPMLIEINSSPTMHPSTPVTQAQLC 215
Db 304 IGDVFEAAATQGIHFQPLENCFEIFGVDFLVDCEQSQVILEVNS-----YPDFKQTGKNL 359
QY 216 AQVOEDTIKVAVD 228
Db 360 SNIENLFSVAVE 372

RESULT 10
S48261
hypothetical protein YBR094w - yeast (Saccharomyces cerevisiae)
N/Alternate names: hypothetical protein YBR0821
C/Species: Saccharomyces cerevisiae
C/Date: 01-Aug-1995 #sequence_revision 11-Aug-1995 #text_change 09-Jul-2004
C/Accession: S48261; S45962; S41800; S44676
R;/Mannhaupt, G.; Stucka, R.; Ehnlé, S.; Vetter, I.; Feldmann, H.
Yeast 10, 1363-1381, 1994
A/Title: Analysis of a 70 kb region on the right arm of yeast chromosome II.
A/Reference number: S48255; MUID:95208357; PMID:7900426
A/Accession: S48261
A/Status: nucleic acid sequence not shown
A/Molecule type: DNA
A/Residues: 1-753 <MAN>
A/Cross-references: UNIPROT:P38254; UNIPARC:UPI000013A3DD; EMBL:X78993; NID:g476045; PID
R;/Feldmann, H.; Mannhaupt, G.; Schwarzlöse, C.; Vetter, I.
submitted to the Protein Sequence Database, August 1994

A/Reference number: S45927
A/Accession: S45962
A/Molecule type: DNA
A/Residues: 1-753 <PE2>
R;/Cross-references: UNIPARC:UPI000013A3DD; EMBL:X35963; NID:g536366; PIDN:CAA85047.1; P
R;/Dekker, P.J.T.; Hoekert, W.; van Oosterum, K.; Grivelli, L.A.
submitted to the EMBL Data Library, December 1992
A/Reference number: S31313
A/Accession: S41800
A/Molecule type: DNA
A/Residues: 167-351, 'TPE', 356-449, 'A', 451-562, 'R', 564-753 <DEK>
A/Cross-references: UNIPARC:UPI0000168D37; EMBL:X69881; NID:g4086; PIDN:CAA49508.1; PID
C/Genetics:
A/Cross-references: SGD:S0000298
A/Map position: 2R

Query Match 11.1%; Score 144; DB 2; Length 753;
Best Local Similarity 21.8%; Pred. No. 4.1e-05;
Matches 57; Conservative 44; Mismatches 105; Indels 56; Gaps 9;

QY 8 WIIPKPAKSRGRDIVCMRVEIE-----LELAAADHP----- 39
Db 476 WIVKPSMSDKGQIRVFKTIEDLQAI FDSFDDSEASESGNDDDDADDVNGEPMDNKNVN 535
QY 40 LGRDNKWVYQKYIETPLLIC---DTKFDIRQWFLVTDNMPLTIWFKYESYLRFSTQRF-- 94
Db 536 ISQLRHFIIQEVLTNPLLLASMDNRKEHRCY--VVCRGDLQVVFVDRMLALFAAKPFVP 593
QY 95 -----SLDKLDSA IHLCNNAVQKYLKNDVGRSPLLPANHNMWTSRFO--EYLQRG 145
Db 594 LDPYAYSVTDLKDLCHLNTCLQS--KKDKDSSVL-----EFDSEIERPNERKS 642
QY 146 AWGSVIYPSMKKAIAHAMKVAQDHVEPRKNSFELYGADPVLGRDRFRPMLIEINSSPTMH 205
Db 643 NI-KEQIHSITNDVFLAAVNAVRLNFQPLNPFYGVDFLIDSNVEYVKLLEINAFPDFK 701
QY 206 PSTFVTAQLCAQVOEDTIKVAV 227
Db 702 QTGKDLKNLIDELFDDTVKYCV 723

RESULT 11
T01270
hypothetical protein At2g19220 [imported] - Arabidopsis thaliana
N/Alternate names: hypothetical protein F27F23.2; hypothetical protein T20K24.23
C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 09-Jul-2004
C/Accession: T01270; T00543; A84574
R;/Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kau
submitted to the EMBL Data Library, May 1998
A/Description: Arabidopsis thaliana chromosome II BAC F27F23 genomic sequence.
A/Reference number: Z14177
A/Accession: T01270
A/Status: translated from GB/EMBL/DDBJ
A/Molecule type: DNA
A/Residues: 1-439 <ROU>
A/Cross-references: UNIPROT:O65915; UNIPARC:UPI00000A9775; EMBL:AC003058; NID:g3135250;
R;/Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes,
submitted to the EMBL Data Library, July 1997
A/Description: Arabidopsis thaliana chromosome II BAC T20K24 genomic sequence.
A/Reference number: Z14167
A/Accession: T00543
A/Status: translated from GB/EMBL/DDBJ
A/Molecule type: DNA
A/Residues: 1-439 <ROW>
A/Cross-references: UNIPARC:UPI00000A9775; EMBL:AC002392; NID:g3176701; PID:g3176720
R;/Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
Muss, D.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,
Nature 402, 761-768, 1999
A/Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

•

RESULT 15

Search completed: April 4, 2006, 12:51:55
Job time : 10.9963 secs

Search completed: April 4, 2006, 12:51:55
Job time : 10.9963 secs